

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 7, 2003, 16:25:31 ; Search time 81 Seconds  
(without alignments)  
1778.110 Million cell updates/sec

Title: US-09-809-638-2

Perfect score: 3720  
Sequence: 1 MTSIMREILLESILGCVSWS.....EGHNYENNHFHMTPKYFL 699

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Sequenced: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_21:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3713	99.8	699	4 Q9H720	Q9H720 homo sapien
2	3125	84.0	699	11 Q91YL7	Q91YL7 mus musculu
3	671.5	18.1	714	5 Q8TIC4	Q8TIC4 dictyostell
4	639	17.2	425	3 Q9P6M4	Q9P6M4 schizosacch
5	625.5	16.8	401	3 P78781	P78781 schizosacch
6	256	6.9	688	3 Q9HD22	Q9HD22 schizosacch
7	141.5	3.8	279	16 Q9S1Z5	Q9S1Z5 streptomyce
8	133	3.6	466	16 Q8Z4F6	Q8Z4F6 salmoneilla
9	132	3.5	583	8 Q950T2	Q950T2 hyaloraphid
10	129	3.5	659	6 Q8SPK7	Q8SPK7 sus scrofa
11	128.5	3.5	396	2 P95556	P95556 pseudomonas
12	128	3.4	466	16 Q8ZMMO	Q8ZMMO salmoneilla
13	124.5	3.3	487	16 Q8ZRI5	Q8ZRI5 salmoneilla
14	124	3.3	609	8 Q79569	Q79569 mustelus ma
15	123.5	3.3	269	11 Q91VY8	Q91VY8 mus musculu
16	122	3.3	636	6 Q9GJX6	Q9GJX6 sus scrofa

17	122	3.3	643	6 Q9GJX7	Q9GJX7 sus scrofa
18	122	3.3	766	17 Q9UYP5	Q9UYP5 pyrococcus
19	121.5	3.3	591	8 Q36149	Q36149 trachymys s
20	121	3.3	605	8 Q9B6V4	Q9B6V4 casuarinus b
21	119	3.2	427	16 Q53515	Q53515 mycobacteri
22	119	3.2	612	8 Q9G6Q3	Q9G6Q3 diplophos t
23	118.5	3.2	449	16 Q9RL05	Q9RL05 streptomyce
24	118.5	3.2	605	8 Q95720	Q95720 casuarinus c
25	118.5	3.2	613	8 Q94SY9	Q94SY9 cretinimugil
26	118.5	3.2	694	16 Q9L179	Q9L179 streptomyce
27	118	3.2	431	16 Q8ZC80	Q8ZC80 yersinia pe
28	117.5	3.2	501	16 Q8Z8J3	Q8Z8J3 salmoneilla
29	117.5	3.2	583	5 Q17069	Q17069 caenorhabdi
30	117	3.1	559	16 Q8Z8E4	Q8Z8E4 salmoneilla
31	117	3.1	604	8 Q9B6Z4	Q9B6Z4 apteryx man
32	116.5	3.1	570	8 Q9B510	Q9B510 tetradoncop
33	116.5	3.1	613	8 Q94S29	Q94S29 mugil cepha
34	116	3.1	557	16 Q8X9F8	Q8X9F8 escherichia
35	116	3.1	605	8 Q9B6W0	Q9B6W0 dinornis gi
36	116	3.1	605	8 Q9B6U8	Q9B6U8 dromaius no
37	116	3.1	605	8 Q958A0	Q958A0 dromaius no
38	116	3.1	958	16 Q8ZFM4	Q8ZFM4 yersinia pe
39	115.5	3.1	590	17 Q96XC0	Q96XC0 sulfolobus
40	115	3.1	317	16 Q97K74	Q97K74 clostridium
41	115	3.1	603	8 Q9MR50	Q9MR50 ciconia cic
42	115	3.1	603	8 Q958D4	Q958D4 tinamus maj
43	114.5	3.1	603	8 Q9GCA20	Q9GCA20 tupia belia
44	114.5	3.1	829	10 Q9LMJ1	Q9LMJ1 arabislopsis
45	114	3.1	395	2 Q9Z404	Q9Z404 pseudomonas

## ALIGNMENTS

RESULT 1

Q9H720 PRELIMINARY: PRT; 699 AA.

AC Q9H720;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DE CDNA: FLJ21511 fis, clone COL05748.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=COLON;  
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,  
RA Oktiani R., Ota T., Suzuki Y., Odayashi M., Nishi T., Shibahara T.,  
RA Tanaka T., Nakamura Y., Isonaga T., Sugano S.,  
RT "NEBO human cDNA sequencing project."  
RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AK025164; BAB15080.1;  
SQ SEQUENCE 699 AA; 78565 MW; 2F839563189A0523 CRC64;

Query Match 99.8%; Score 3713; DB 4; Length 699;  
Best Local Similarity 99.9%; Pred. No. 3, 6e-263;  
Matches 698; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTSIMREILLESILGCVSWSLYHDLPMTYYPFLQTLTGEGFSIAFLSPFLITTF 60  
DB 1 MTSIMREILLESILGCVSWSLYHDLPMTYYPFLQTLTGEGFSIAFLSPFLITTF 60

QY 61 WKLWNKMWLTLLRITIGSIASFOAPNAKRLRLMALGVSSLIYQAVYTWMSGSLQRY 120  
DB 61 WKLWNKMWLTLLRITIGSIASFOAPNAKRLRLMALGVSSLIYQAVYTWMSGSLQRY 120

QY 121 LRIWGFILGOIVLVIRIWTSLNPITWSQMSKNVLTLSAINTLDRIGTDGCSPEEK 180  
DB 121 LRIWGFILGOIVLVIRIWTSLNPITWSQMSKNVLTLSAINTLDRIGTDGCSPEEK 180

OY	181	KYGEVATGASPNMLLAGAFCGLVPLTTHVFEBSVLSWMAVSGHPHGPONPBGCA	2400
Db	181	KTGEVATGASBPNNKLLAGAAFCGLVPLTTHVFEBSVLSWMAVSGHPHGPONPBGCA	2400
OY	241	VLLCLASGLMLPSCILMFRCGLIMWYGTGTASAAGILYLTHTMAAVSGCVAIFITASMPQ	3000
Db	241	VLLCLASGLMLPSCILMFRCGTGLIMWYGTGTASAAGILYLTHTMAAVSGCVAIFITASMPQ	3000
OY	301	TLGHILNGSTNPGKMTJAMTFYLLTPEFCMCAFPFVPGVYARERSDVLGTMMILII	3600
Db	301	TLGHILNGSTNPGKMTJAMTFYLLTPEFCMCAFPFVPGVYARERSDVLGTMMILII	3600
OY	361	GLNMLFGRKMLDLLOTKNNSKYLFRKSEKMYMLFTKMLLVGVLGLGLGRHKAAYERKLG	4200
Db	361	GLNMLFGRKMLDLLOTKNNSKYLFRKSEKMYMLFTKMLLVGVLGLGLGRHKAAYERKLG	4200
OY	421	KVAPPKVEYSAIIPWPRRGYDNEGSSSLERSAHLNENFGADFTITLESDAKPYMGNDLT	4800
Db	421	KVAPPKVEYSAAIIPWPRRGYDNEGSSSLERSAHLNENFGADFTITLESDAKPYMGNDLT	4800
OY	481	IMWLGEKLFYTFDFGSTRYHTMGJMALSRYPYKSEHNLHPSPGCEIAPAPITTLTVNISK	5400
Db	481	IMWLGEKLFYTFDFGSTRYHTMGJMALSRYPYKSEHNLHPSPGCEIAPAPITTLTVNISK	5400
OY	541	LYDEVYTHFGHNEDDLDRKLOAIAVSKLLKSSNOVIFLGYTTSAPGSRDYLOLTHEGNV	6000
Db	541	LYDEVYTHFGHNEDDLDRKLOAIAVSKLLKSSNOVIFLGYTTSAPGSRDYLOLTHEGNV	6000
OY	601	KDIDSTHDMRCEYIYMYGGLRLGYARSHAELSDEIOAKKPIIPDDPNTYNDONKYYI	6600
Db	601	KDIDSTHDMRCEYIYMYGGLRLGYARSHAELSDEIOAKKPIIPDDPNTYNDONKYYI	6600
OY	661	DHREYSEKIHNPREPGYKESGANTENNHHFHMANTPKYFL 699	
Db	661	DHREYSEKIHNPREGYSKEGHANTENNHNENFMANTPKYFL 699	

D	b	121	LKINGFLIGHVILLVLRITWTSILNPIMSYQMSNRVILITLSAVADRIGDGYRNPCK	180
Q	y	181	KTEGVAATGMSAPRWMLLAGAAGSVLFETHWVGEVSVSRMAVSGAPHPGDPNPPGA	240
D	b	181	KPREVAAGRTSLSSMLLPGLGAFGOSLFLFTWIFGEVSVSRMAVSGAPHPGDPNPPGGA	240
Q	y	241	VLLCAGSLMIPSCITMFRGGLIMWVGTASAAGLYLPHWMAAVSGCAVIFTAMMP	300
D	b	241	VLLCFSSGSLMISGSHLHDGAGLAMMMGAASAGLLYLRHMAAVSGCVLAVFTGSMRP	300
Q	y	301	TLGHLINSSTNPCKTMTIAMFYLLIEFCAMCTAEKREYGVYARESDVLLGTMMII	360
D	b	301	VLGHLVNSGKSGCEAMATGMLLYLOFFCAMCTAEKREYGVYARESDVLLGTIMVII	360
Q	y	361	GLNMLFPPKKNLDLLIOTKUSSKYLFPKSKKMKLFMLLYGVGLIGLIRHAYBERKJ	420
D	b	361	GLSMLFPPKKNLDLLOTKUSSPPTLLECSSEKMKLIMLELVGVGLLGLIRHAYBERKJ	420
Q	y	421	KVATKEVSAAIWPFRRGYDNEGSSLEERSAHLINTGADFIITLESDAKPYMGNNDLT	480
D	b	421	RGAPATVYSAAIWPFRRGYDNEGPNLEBSAOLLEKTAGDFIITLESDAKPYIGNNDLT	480
Q	y	481	MLLCEKICGFITTDGPRTRYTWGIMALSRIPIYKSEHNLPSPEGELAPITLTVINSG	540
D	b	481	MLLCEKICGFITTDGPRTRYTWGIMLSRPIYKSEHNLPSPEGELAPITLTVINSNR	540
Q	y	541	LVDPEVYHFGNHEDDRLQALAVSKLSSNOVYFLGYIPSAGRSDYLDLTGHCNV	600
D	b	541	LVDPEVYHFGNHEDDRLQALAVSKLNCNOVYFLGYIPSPGRSDYLDLTGHCNV	600
Q	y	601	KDIDSTHDKWCEYIMRGILRLGAYARIASHLESDSEIQAKRRIDDPNTRYNDKNVYI	660
D	b	601	KDIDSSGDGRCEYIMRGILRLGAYARIASHLESDSEIQAKRRIDDPNTRYNDKNVYI	660
Q	y	661	DHREVESEKIHNPFRGSGYKGHVYENNNHNNMTPKYFL	699

RESULT 2  
Q91YL7

ID	091YL7:	PRELIMINARY:	PRT:	699 AA.
DT	01-DEC-2001 (TREMBLrel. 19)	Created		
DT	01-DEC-2001 (TREMBLrel. 19)	Last sequence update		
DT	01-MAR-2002 (TREMBLrel. 20)	Last annotation update		
DE	Similar to hypothetical protein FLJ21511.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus			
OX	NCBI_TaxID=10090;			
RH	[1]			
RE	SEQUENCE FROM N.A.			
RF	Strausberg R.;			
RL	Submitted (Oct-2001) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: BC016523; AAHL552.1; "-			
DR	InterPro: IPR001092; HLM basic.			
DR	PROSITE: P500038; HELIX_LOOP_HELIX; UNKNOWN_1.			
KW	Hypothetical protein.			
SC	SEQUENCE 699 AA: 78179 MW: 4F41975D6570D5F8 CRC64:			

Query Match	84.0%	Score 3125	DB 11	Length 659
Best Local Similarity	82.8%	Pred. No. 3.7e-220		
Matches 579; Conservative	41	Mismatches 79	Indels 0	Gaps 0

[illegible]

### RESULT 3

Q8BTIC4	PRELIMINARY;	PRT;	714 AA.
ID Q8BTIC4			
AC Q8BTIC4			
DT 01-JUN-2002 (TrEMBLrel. 21, Created)			
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)			
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)			
DE Hypothetical 79.7 kDa protein.			
OS Dictyostelium discoidium (Slime mold).			
OC Eukaryota; Mycetozoa; Dictyostelidida; Dictyostelium.			
OX NCBI_TaxID=44689;			
RN [1]			
RP SEQUENCE FROM N.A.			
RC STRAIN-AA4;			
RA Goeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,			
RA Lehmann R., Baumgart C., Parra G., April J.F., Guiso R., Kumpf K.,			
RA Tungal B., Cox E., Quail M.A., Platzter M., Rosenthal A., Noegel A.A.			
RT "Sequence and Analysis of Chromosome 2 of Dictyostelium."			
RL Submitted (Apr-2002) to the EMBL/Genbank/DBU databases.			
RL EMBL: AC116960; AAM08475.1; -			

KW Hypothetical protein.  
SQ SEQUENCE 714 AA; 79745 MW; 82F773B4CC75C349 CRC64;

	Query Match	18.1%	Score 671.5;	DB 5;	Length 714;	
	Best Local Similarity	27.1%;	Pred. No.	1.1e-40;		
	Matches 188;	Conservative 147;	Mismatches 281;	Indels 79;	Gaps 22;	
QY	8 ILLESLGCWSMLVHDCPMIYYPLDQLTLETGEFSIAFLPFIETTPTEPWKLUNK - 66	:	:	:	:	:
	: : : :	:	:	:	:	:
Dd	25 LLDAGLIFAYNNLSLMEVFISVIWPFMPMGFGCTGEAALLITFPILFGLGITNRIYSRF 84	:	:	:	:	:
OY	67 -----KMKLTLLRLIRTT-----GSIAS-----EPAPNAKLRLLWLALGVSSLIYAQVT 110	:	:	:	:	:
Dd	85 NLPTRIVAOULLAFPAIFPAIDKASIKDKDEGISSPIVV-KTFEVSVAVALDLMDLAOCNS 143	:	:	:	:	:



```

Db 180 LIMGFRDLTQVLAHDLGADYGPDPKHTWGAALLSKFPIVNSTHLLPSPQCELAFAI 239
QY 532 TLTVNIGSKLVDFTVTHGHNHEDDLDRKLAIVANSKLLSSNOVIFLGYTSPGSDY 591
Db 240 HATLDVYGLIDVYVSHNGOESQDLRLOSTELARRESPLRGVYVSNVGOEPO 299
QY 592 LQTEHGNVNDIDSTDHDMCEYIMYRGLIRAGYARISHAELSDSEIQMAKFRIPDPPTN 651
Db 300 TILTRDGMMDIEPADYDRMCQYIFRGVYKIRIGYARLHRSITTELOCTGKFLYTKDL- 357
QY 652 YRDQKAVYIDHREYSEKIHENPRF-GSYKECHYENNNHFMNTPKYF 698
Db 358 ---GRNVRIKDEHYESHRYPSLEGTGVNCHYDNLVH--BPWY 400

RESULT 6
ID Q9HDZ2 PRELIMINARY; PRT; 688 AA.
AC Q9HDZ2;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
D 01-MAR-2001 (Tremblrel. 16, Last annotation update)
G Hypothetical 77.5 kDa protein (Fragment).
GN SPAC589.12.
OS Schizosaccharomyces pombe (Fission Yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes;
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972H-;
RA Wood V., Rajandream M.A., Barrell B.G., Aert R., Robben J.,
RA Meljens I., Grymoprez B., Volckaert G.,
RA Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
RL EMBL: AL512496; CAC19769.1;
RM Hypothetical protein.
FT NON_FER 688
SQ SEQUENCE 688 AA; 77474 MW; 3985FAC23CDA34FF CRC64;

Query Match 6.9%; Score 256; DB 3; Length 688;
Best Local Similarity 22.5%; Pred. No. 2.6e-10;
Matches 97; Conservative 77; Mismatches 197; Indels 60; Gaps 11;
QY 11 ESLLGCVSWSLYHDLGMIYFPLQTELEGLCEFSIAFLSPILITTPFKLVNKKMML 70
Db 285 EYVLSFVFWSLVSLGLLWTFPLMHGICSGYECILFELSPLLGIPLLRKFAKVPVI 344
QY 71 TLRTITIGSIASFQADPNALRLMVLALGVSSSLIVQAVTW-----SGSHLQRYLRIV 124
D 345 FL--FLNVITGIAAKKLEDPVHRLEFVTA-----SYCCCELANTSIFSINSPENLAIEKIS 398.
QY 125 GFLIGQIVLVLRITWYSLNFIW-----SYQMSNRKVIITLSAIVTLDRIGTGDGCSK 176
Db 399 TFLGGLLASSIAKYSFSPNNPIMFILNETNGKQIPALIVIGIICLFAIFPHVOQTANA 458
QY 177 PEEKKTGEVATGASRRNMLLAGAFSGSLVFLTHWVFGVEVSLVSRNAVSGHPRHGPDPNP 236
Db 459 VEHRKLKRTA-----LSAALSLGTVLFCLTFLCDSTVMTWMSWDGPIRGPOYP 510
QY 237 FGAVALCLASGLMPSCLMFRGTGLIMWTGTASAGLLYL--HTWAAVSGCFAEFT- 294
Db 511 HGAIVSYVSCAVAVAYRVLQSGAFMLIGFVLCFSGSYEMTINHGKCSYLGGLIFTSYVL 570
QY 295 ---ASMPQTLGLHNSGTPGKTMTIAMIFYLLEIFPCAMCTAFKFEVGVYARERSD 350
Db 571 IYSPASI-----RISSEYSPAKVWGCAFLVILYSLAHVWVAVYFVPGCPILBERTS 623
QY 351 VLCTMMLIIGLNLF-----GPKKNDLLIOTKSSKVLFRKSEKYMFLMML 400
Db 624 YIL---IFIGMNLALVPAVSGESKEPNKADSSVVDIKOSDSSRRSRSPFKS-----LL 674

```

```

QY 401 VGVGLIGLR 411
Db 675 TGCELAIMALK 685

RESULT 7
ID Q9S125 PRELIMINARY; PRT; 279 AA.
AC Q9S125;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
D Putative secreted protein.
GN SC00131 OR SCJ21.12C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Seeger K.J., Harris D.,
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; Pubmed=8843436;
RA Redenbach M., Kleser H.M., Denapalte D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome."
RL Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2)."
RL Nature 417:141-147(2002).
RL EMBL: AL109747; CAB52358.1;
DR InterPro: IPR005135; Exo_endo_phos.
DR Pfam: PF03372; Exo_endo_phos.1.
SQ SEQUENCE 279 AA; 30031 MW; B16719A1698580C0 CRC64;

Query Match 3.8%; Score 141.5; DB 16; Length 279;
Best Local Similarity 24.9%; Pred. No. 0.019;
Matches 63; Conservative 36; Mismatches 91; Indels 63; Gaps 12;
QY 447 LERSAHLNLTGGADFTTILSDASKPYMGNDLJMW-----LGEKLGFTDGE----- 494
Db 58 LRRYAVTRKSGADVGLQVD--KHSARSD--WADQPAELAEELGHVVEGANINDS 112
QY 495 ---PSTRYHTWGLMSLRYPYVSEH--HLPSPEGEIAPAITLVNISKLYDFTVTHF- 549
Db 113 PPAPEGHRVGYGALLSRYPITASDNTWLYKSPGQEGRLHMTLDVHGKKEVEFYMTHLA 112
QY 550 -GNHEDDLRKLQAIIVANSKLLSSNOVIFLGYTSPGSDYDLQLTEHGNVNDIDSTDH 608
Db 173 AGSQADRLQDTAQQVVDLIGTRPG---LLVGDFNALPAPE-----SRPLQNAVY 219

```



OY 329 FCAMCTAFKFPGVYARERSDV-----LGTMLLIGLMLFGPKMLDLLQT 378  
 DB 273 ILIWLGAATAFGAVYGYVOTDLKRTIAYSTSTSLGYMLVACLG----- 317  
 OY 379 KNSKYLFRSEKRYKMLFLLVGVGLGLGNKAYE-----KRLGVAPFKEY----- 428  
 DB 318 QYGLALHLVNAHFAFLALSLAG-----SVLHAIHDEODIRKMGIMSPVYITSM 371  
 OY 429 -----SAAIMPFREGYDNEG-----SLSERSAHLNETGADFTILES-----DAS 470  
 DB 372 VGSLSLVALPEFLGYYSKIDILQTAECTGYLNG-LGAATFATFYSLKLHRYFWLPPQS 430  
 OY 471 KPYMGNDLTMWLGKLGFTYDFGPTRYHTWGMALSRPIVSEHLLPSPEGEI 527  
 DB 431 KVLGAHEPSAMMLIPTTLTLEFSIS-----MGYLAQNHAAALAPAMNTLFLPQNI 482

## RESULT 10

O8SPK7 PRELIMINARY; PRT; 659 AA.  
 AC O8SPK7:  
 DT 01-JUN-2002 (Tremblrel. 21, Created)  
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE Sodium iodide symporter.  
 GN NIS.  
 OS *Sus scrofa* (Pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 OX NCBI\_TaxID=9823;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=THYROID.  
 RA Selmi-Ruby S., Roussel B.;  
 RT "Molecular cloning and functional analyses of pig sodium iodide  
 RT symporter: evidence for three forms generated by alternative  
 RT splicing."  
 RL Submitted (May-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AJ277989; CAC81949.1;  
 SO SEQUENCE 659 AA; 70125 MW; 3C06C8691435E68 CRC64;

Query Match 3.5%; Score 129; DB 6; Length 659;  
 Best local Similarity 21.9%; Pred. No. 0.48;  
 Matches 68; Conservative 42; Mismatches 107; Indels 94; Gaps 14;

OY 15 GCYSMSLYHDLGPMI-----YFPLQTL-----ELTGLEFSIAFLSPITLTPFMK 62  
 DB 300 GIWFAFLYDCDPLAGHSAPDOYMPLLVLDIFEDLPVPG-----LFLACAVSGT 351  
 OY 63 LVNKKMMLTLRITITIGSIASFOAPN-AKLRLVNLVAGVSSLIYQAVTWMSGSHLQRYL 121  
 DB 352 LSTVASTSINMAAVYEDLDIKPRLNAPRLVITISGLSCLVALSLSGGVLAGSF 411  
 OY 122 RINGFI-----LGOIVL-----VLRIRWY-----TSLNPINSTYQNSNKY 155  
 DB 412 TVMGVLSIGPLGFAFVGMFLPSCNTSGVLSGLAAGLALSLVAVGASLYPPSAQSAGS--- 468  
 OY 156 ILTFLSAIATLDRIGTDGCKPEEKKTG-----EVATGMSRPMWLLAGAFGS 204  
 DB 469 VLPSSA-----AGCLLPTANASGLDDPYLAVANASTASSLEDPDPLIASFYA 518  
 OY 205 LVFLTHWVEGVSLSRMAVSGHPHGPDPNPGAVLLCLASGLMLPCLMFRGTGLIW 264  
 DB 519 ISLYLYGALGTSLT-----LCGALISCL-TGPIKRSAL-----GPGLLW 557

OY 265 W--VTGTASA 273  
 DB 558 WDLTROTAASYA 568

## RESULT 11

p95556

ID P95556 PRELIMINARY; PRT; 396 AA.  
 AC P95556:  
 DT 01-MAY-1997 (Tremblrel. 03, Created)  
 DT 01-MAY-1997 (Tremblrel. 03, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE ORF396 protein.  
 GN ORF396.  
 OS *Pseudomonas stutzeri* (*Pseudomonas perfectomarina*).  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 OX NCBI\_TaxID=316;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ZOBELL ATCC 14405;  
 RX MEDLINE=97107629; PubMed=8950369;  
 RA Glockner A.B., Zunft W.G.;  
 RT "Sequence and analysis of an internal 9.7-kb segment from the 30-kb  
 RT denitrification gene cluster of *Pseudomonas stutzeri*."  
 RL Biochim. Biophys. Acta 1277:6-12(1996).  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ZOBELL ATCC 14405;  
 RX MEDLINE=96028114; PubMed=7588711;  
 RA Palmedo G., Seither P., Koerner H., Matthews J.C., Burkhalter R.S.,  
 RA Timkovich R., Zunft W.G.;  
 RT "Resolution of the nrd locus for heme d1 synthesis of cytochrome cdl  
 RT Eur. J. Biochem. 232:737-746(1995).  
 DR EMBL: Z73914; CAA98157.1;  
 SO SEQUENCE 396 AA; 43083 MW; EB07255ECFB1E0D3 CRC64;

Query Match 3.5%; Score 128.5; DB 2; Length 396;  
 Best local Similarity 21.1%; Pred. No. 0.27;  
 Matches 82; Conservative 55; Mismatches 147; Indels 105; Gaps 20;

OY 55 LITPPKLVNKKMML--TLRITIT-----GSIASFOAPNAKL-----RLWVLAL 98  
 DB 10 LSTAPITRLAFRPFPLAGSLYALAIPLVVAWMTGLMPGFGWGLAMHREMLFGFAM 69  
 OY 99 GVSSSLIYQAVTW-----SGSHLQRYLRW-----GFLG-----QIVLVLR 338  
 DB 70 AIVAGFLTRAVOTGTOTAPSGNRLVGLAAVLAARLGLFGLPAAMLAFLDLFLVALV 129  
 OY 139 WYSLNPIMSYQMSNK--VILTLSAITATLDRIGTDGCKPEEKKTGEVATGMASRPM 196  
 DB 130 WMAA-QMLNAVROKRNYPVIVVLSLMLGADVLLTLLGNDALQOGVLAG-----WL 183  
 OY 197 LAG--AAGSLV--FLTH-----WVEGVSLSRMAVSGHPH--GDPNPG 238  
 DB 184 VAALMLLIGGRVLPFTORGKVDADVKKPMWLDVALLGVIALLHAFGVAMRPOP 243  
 OY 239 GAVLLCLASGLMLPCLWF--RG-----TGLIW-----WVTGTAAGLYLHTMAAVS 286  
 DB 244 GLLFVAIGVGHILRLKRWTDKIGKVGLLSLVAMLM--VVAAGGLAMHFGLLAQS 301  
 OY 287 -----GCVFAIFTASMPQTLGH-----LINSITNPGRKTMITAMIF 322  
 DB 302 PSLHALSVGSMGLILAMIAVTLTGTRPDLQPAIGAFVLFNIGTARVFLSVAMPV 361  
 OY 323 YLEIFFCAMCTAF-----KVPDGYAR 346  
 DB 362 GGLMAVACWTLAFALYVWRVAPMLVAAR 390

## RESULT 12

O8ZMKO PRELIMINARY; PRT; 466 AA.  
 AC O8ZMKO:  
 DT 01-MAR-2002 (Tremblrel. 20, Created)  
 DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE APC family, gamma-aminobutyrate transport protein, Rpos  
 dependent.

RESULT 13		
08ZRI5		
08ZRI5	PRELIMINARY;	PRT; 487 AA.
AC 08ZRI5:		
DT 01-MAR-2002	(TREMblrel. 20, Created)	
DT 01-MAR-2002	(TREMblrel. 20, Last sequence update)	

RESULT 14	
079569	
ID 079569	PRELIMINARY; PRT; 609 AA.
AC 079569;	
DT 01-NOV-1998 (TREMblrel, 08, Created)	
DT 01-NOV-1998 (TREMblrel, 08, Last sequence update)	
DT 01-MAR-2002 (TREMblrel, 08, Last annotation update)	
DE NADH dehydrogenase subunit 5.	
CN NADH5	
OS Mustelus manazo.	
OG Mitochondrion.	
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;	
OC Elasmobranchii; Galeomorphi; Galeoidei; Carcharhiniformes; Triakidae	
OC Mustelus.	

NCBI\_TaxID=79736;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=LIVER.  
 RX MEDLINE=99083431; PubMed=9866199;  
 RA Cao Y., Maddell P.J., Okada N., Hasegawa M.;  
 RT "The complete mitochondrial DNA sequence of the shark (Mustelus  
 manaco): Evaluating rooting contradictions to living bony  
 vertebrates.";  
 RL Mol. Biol. Evol. 15:1637-1646(1998).  
 CC -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.  
 DR EMBL: AB015962; BA033045.1;  
 DR InterPro: IPR003916; NADhub\_oxred5.  
 DR InterPro: IPR001750; Oxidored\_q1.  
 DR InterPro: IPR001516; Oxidored\_q1\_N.  
 DR InterPro: IPR000215; Serpin.  
 DR Pfam: PF00361; oxidored\_q1.  
 DR Pfam: PF00662; oxidored\_q1\_N.1.  
 DR PRINTS: PR01434; NADHDHGNSE5.  
 DR PROSITE: PS00284; SERPIN; UNKNOWN\_1.  
 DR MitoChondrion; NAD; Oxidoreductase; Ubiquinone.  
 KW SEQUENCE 609 AA; 68215 MW; 97FBE6D34CDE7DD0 CRC64;

Query Match 3.3%; Score 124; DB 8; Length 609;  
 Best Local Similarity 19.3%; Pred. No. 1;  
 Matches 131; Conservative 83; Mismatches 210; Indels 254; Gaps 32;

155 VILTLAATLDRICTDSCSKPEKKTGEVATGMSRPNM--LLAGAA-----FGSLV- 206  
 Db 14 IILPLISSL-----SRKELK-----PNNSSLYVTAVKISFISLIP 52  
 207 -----FLTHW-----VFGEVSLVSRMAY-----SGHPHP 230  
 Db 53 LEFLDGLSEIVYNNMMNMPDINMSKFDLYSIIFPVALYVMSLIEFALWMHS 112  
 231 GPPNPFGGAVLLCLAGLMLPSC-----LMFRGTGLI-----WVVT-----GTASAA 273  
 Db 113 DPMNNREFKYLLEFLISMILVYANNMFOLFEGEIVMSFLIGWYSRADANTALQ 172  
 274 GLLY-----LHTW-----AAVSGCVFAI 292  
 Db 173 AVIYNRGDGLISMAMLTNLMSWEIHOLFISKNKDLTLPLGLVLAAGSAOGL 232  
 293 F-----TASMPQTLGHLINSNTNPGKTMITAMFYLLER-----FCAMCTAFK 337  
 Db 233 HPMLPSAMEGPTFVSALLHS-----STMVAGIFLLIRLPLQDNKLLITVCLGALT 287  
 338 FVYGGVYARERSVY-----LGTMLLIGLNNLFGPKKNLDDLLOTKSSKYLFR 387  
 Db 288 TLEFATCALTONDIKIVASTSSQGLMVTIGLN-----QPOLAFLHI 332  
 388 KSEKYNKLFMLLVGVGLGLGRHKAEE-----RLG-----KVAPTKEVSAIIMPFREGYD 440  
 Db 333 CTHAFKAMFLCSG-----SIHSLNDODIRKMGKGLKLPFTSTLTIGS----- 380  
 441 NEGMSSLESAHLINLETGADFIT--IESDASKPYMGNNDLTMMLGKGLGYTDFGPESTRY 499  
 Db 381 -----LALTMGFLSGFSGFKDAIIESMNTSHLNAWALLITVATSP---TAI 424  
 500 HTWGM--ALSRYP-----IVKSEHLLPSP-----EGEIAAIVITVNI SGKLVDFV 546  
 Db 425 YSLRLIFALMNPREFNTLSPINENNPLVINPIKRLAYGSIAGLITLMLT----- 476  
 547 THFGNEHDDDRKLOATAVSKLSSNOVIIFGIYTSAGSRDYLOUTHEGNVKIDIST 606  
 Db 477 -----PTKQITMTSPPLKLSALLVITINGLLA-----LETLNLTN----- 512  
 607 DHDRCCEYIYWRGLIR-LGY-ARISHAELSDSEIOMAKF---RIPDDPTVYRDNOKYVID 661  
 Db 513 SHFKTNPETLHYHHFNSNLGYPFSIIRHLPKTSLNMAQYISTHLLDQYWEKIGPKSNLI 572  
 662 HREVSKEIHNNPFGSKY 679

Db 573 QQTSLIKLSTPOQGLIK 590

## RESULT 15

ID Q91YV8 PRELIMINARY; PRT: 269 AA.  
 AC Q91YV8;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Aquaporin 1.  
 GN AQP1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-BREAST TUMOR;  
 RA Strausberg R.;  
 RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: BC007125; AAH07125.1;  
 DR MGD: MGI:103201; Agpl.  
 DR InterPro: IPR000425; MIP\_family.  
 DR Pfam: PF00230; MIP; 1.  
 DR ProDom: PD000295; MIP\_family; 1.  
 DR TIGRFAMs: TIGR00861; MIP; 1.  
 DR PROSITE: PS00221; MIP; UNKNOWN\_1.  
 KW Porin.  
 SO SEQUENCE 269 AA; 28823 MW; 80534DE4B78AB5E7 CRC64;

## Query Match

3.3%; Score 123.5; DB 11; Length 269;  
 Best Local Similarity 22.8%; Pred. No. 0.37;  
 Matches 60; Conservative 36; Mismatches 82; Indels 85; Gaps 11;

60 FKLVLNKKM-LTLRLITIGSIASFOAPNAKURLMAY-----LALGVSSSLIYQAVTW 111  
 Db 10 FMRVVAEFLAMTLPEFISIGSALGFNYPLERNQTLVQDNVKSFLAFGLSIATLAOSVGH 69  
 112 WSGSHLQRYLRINGFILGOIVLVLRIMY-----TSINPLMSYQMSNK----- 154  
 Db 70 ISGAHLNPAVTL-GLLSCQISILRAWMYIIAOCVGAIVATAILSGITSLVNSLGRND 128  
 155 -----VILTLAATLDRICTDSCSKPEKKTGEVATGMSRPNM 196  
 Db 129 LAGVNSGGGLGIEITIGTLQVLVCLATDORRRDLGGSAP-----LAIGSLVALGH 181  
 197 LA-----GAFGSLV-----FLTHVFGVSLVSRMAYSGHPGDPNPF-GGAV 241  
 Db 182 LAIDYGCISINPARSFGSAVLTRNFNSHWIF-----WV-----GPFIGAL 222  
 242 LCLASGLMLPSCLMFRGTGLIW 264  
 Db 223 AVLITDFILAPRSSDPTDRMKW 245

Search completed: January 7, 2003, 22:27:46  
 Job time : 95 secs



124

ALIGNMENTS

RESULT 1  
AK025164  
LOCUS AK025164 2486 bp mRNA linear PRI 29-SEP-2000  
DEFINITION Homo sapiens cDNA: FLJ21511 fis, clone COL05748.  
ACCESSION AK025164  
VERSION AK025164.1 GI:10437625  
KEYWORDS oligo capping; fis (full insert sequence).  
SOURCE Homo sapiens colon cDNA to mRNA, clone\_lib:COL clone:COL05748.  
ORGANISM Homo sapiens

**THIS PAGE BLANK (8719)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 7, 2003, 13:40:54 : Search time 5427 Seconds  
(without alignments)  
11277.538 Million cell updates/sec

Title: US-09-809-638-1

Perfect score: 2103  
Sequence: 1 atgacctgcgtgagaga.....ccaaatactttatgaac 2103

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Sequenced: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_hlg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_hlg\_hum:\*  
31: em\_hlg\_hum:\*  
32: em\_hlg\_hum:\*  
33: em\_hlg\_hum:\*  
34: em\_hlg\_hum:\*  
35: em\_hlg\_hum:\*  
36: em\_hlg\_hum:\*  
37: em\_hlg\_hum:\*  
38: em\_hlg\_hum:\*  
39: em\_hlg\_hum:\*  
40: em\_hlg\_hum:\*  
41: em\_hlg\_hum:\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2098.1	99.8	2486	9 AK025164	AK025164 Homo sapi
2	1511	71.8	2412	10 BC016523	BC016523 Mus muscu
3	267.4	12.7	184206	9 AC020593	AC020593 Homo sapi
4	192.6	9.2	218040	2 AC108828	AC108828 Mus muscu
5	191	9.1	162312	2 AC114455	AC114455 Rattus no
6	160	7.6	222885	2 AL158816	AL158816 Homo sapi
7	151	7.2	222885	2 AL158816	AL158816 Homo sapi
8	132.4	6.3	316613	8 SCCHRIIT	X59720 S.cerevisia
9	122.6	5.8	218040	2 AC108828	AC108828 Mus muscu
10	121.8	5.8	1234	8 D89130	D89130 Schistosac
11	121.8	5.8	34034	8 SPAC688	AL355632 S.pombe
12	112	5.3	305	6 AX337478	AX337478 Sequence
13	110.6	5.3	995	11 CNS06JW	AL402149 T3 end of
14	108	5.1	162312	2 AC114455	AC114455 Rattus no
15	81.6	3.9	170472	2 AC106576	AC106576 Rattus no
16	81.4	3.9	1348	9 HMTWPIPSA	K03223 Human TPI (
17	59.8	2.8	171187	2 AC116960	AC116960 Dictyoste
18	54	2.6	54	6 AX106346	AX106346 Sequence
19	54	2.6	54	6 AX140637	AX140637 Sequence
20	54	2.6	54	6 AX200497	AX200497 Sequence
21	54	2.6	54	6 AX267153	AX267153 Sequence
22	42.6	2.0	178874	2 AC095821	AC095821 Rattus no
23	42.4	2.0	68261	2 AC115084	AC115084 Homo sapi
24	42	2.0	162349	9 AC006226	AC006226 Homo sapi
25	42	2.0	190000	2 AC007072	AC007072 Homo sapi
26	42	2.0	192987	2 AC110450	AC110450 Rattus no
27	41.6	2.0	267698	2 AC123071	AC123071 Mus muscu
28	41.4	2.0	161990	2 AC120482	AC120482 Rattus no
29	41.4	2.0	167750	2 AC016439	AC016439 Homo sapi
30	41.4	2.0	172710	2 AC021928	AC021928 Homo sapi
31	41.2	2.0	141577	9 AL606748	AL606748 Human DNA
32	41	1.9	6563	10 RAYCP450A	M26129 Rat cytochr
33	41	1.9	208805	10 AL627072	AL627072 Mus muscu
34	41	1.9	286358	2 AL713872	AL713872 Mus muscu
35	40.8	1.9	408	11 G51484	G51484 SHGC-82855
36	40.8	1.9	42114	3 U50193	U50193 Caenorhabdi
37	40.8	1.9	110000	2 AC117400-2	Continuation (3 of
38	40.8	1.9	124067	9 AP000435	AP000435 Homo sapi
39	40.8	1.9	147358	9 AP002345	AP002345 Homo sapi
40	40.6	1.9	95146	2 AC099240	AC099240 Rattus no
41	40.6	1.9	161892	2 AC123875	AC123875 Mus muscu
42	40.6	1.9	195790	2 AC124505	AC124505 Mus muscu
43	40.4	1.9	100061	2 AC016551	AC016551 Homo sapi
44	40.4	1.9	170569	2 AC108109	AC108109 Homo sapi
45	40.4	1.9	191132	9 AC010282	AC010282 Homo sapi

## ALIGNMENTS

RESULT 1  
AK025164  
LOCUS AK025164 2486 bp mRNA linear PRI 29-SEP-2000  
DEFINITION Homo sapiens cDNA: FLJ21511 fls, clone COL05748.  
ACCESSION AK025164  
VERSION AK025164.1 GI:10437625  
KEYWORDS oligo capping; fls (full insert sequence).  
SOURCE Homo sapiens colon cDNA to mRNA, clone.lib:COL clone:COL05748.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE  
1 (sites)  
Kawabata,A., Hiki,T., Kobatake,N., Inagaki,H., Ikema,Y.,  
Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,  
Shibahara,T., Tanaka,T., Nakamura,Y., Isegai,T. and Sugano,S.

TITLE NEDO human cDNA sequencing project  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 2486)  
AUTHORS Sugano, S., Suzuki, Y., Ota, T., Obayashi, M., Nishi, T., Isogai, T.,  
Shibahara, T., Tanaka, T. and Nakamura, Y.  
TITLE Direct Submission  
JOURNAL Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science,  
University of Tokyo, Laboratory of Genome Structure Analysis, Human  
Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639,  
Japan (E-mail: cdna@leims.u-tokyo.ac.jp, Tel: 81-3-5449-5286,  
Fax: 81-3-5449-5416)  
COMMENT NEDO human cDNA sequencing project supported by Ministry of  
International Trade and Industry of Japan: cDNA full insert  
sequencing: Research Association for Biotechnology: cDNA library  
construction, 5'- & 3'-end one pass sequencing: Department of  
Virology and Human Genome Center, Institute of Medical Science,  
University of Tokyo (partly supported by Science and Technology  
Agency).

FEATURES  
source

## CDS

1..2486  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="COL05748"  
/tissue\_type="colon"  
/clone\_1ib="COL"  
/note="cloning vector pME18SFL3"  
183..2282  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="BAB15080.1"  
/db\_xref="GI:10437626"  
/translation="MTSLMREILLESLLGCVSMYLDHPMIVYFPLOTLENLGEG  
FSIAFLSPFLFTPEFKLVNKKMLTLLRITTSISFOFNPKNLRMLVALGVSS  
SLIVQVTFMSSSHLORLIRMGFLTIGLIVLIRVMTSLIPMSYMSKNVILITLS  
AIATIDRIGTDGDCSKPREKKTGEVATGNASRPNNLLGAAAGSVPLTHWFGVSL  
VSRMVASGHPHPDPNPBGAVLLCLASGLMPSCLMFRGTGLIMWTGTSAAGL  
LHTMAAVVSGCVFAIFTSAMPOTLGHLSNGTNPGRMTIAMFYLLFETFCAMCT  
AFKPEVGVYARERSDVLGTMLIIGLNLFGPKKNDLLQTRNSKVLFRKSEKY  
KGLFLMLVGVGLGLGRHKAVERKKVAVPTKVSAAIMFRRGSDYNEGMSLERS  
AHLNMGADFTTILSDASPRYNNLMTLGRKLFEPYDPSSTYHNMGLTALS  
RPIYKSEHHLIPSPBEGEAPRLITTVINSKLVDPYVYTHGNEHDDIDRKLQALAYS  
KLKSSSNQVIFLGTTSAPGSRDYLQTLQHEHNVDDIDSTDRNCEYIMRGLRLG  
YARISHAELSDSEIQMAKFRIPDDPTNVRDNRKVVIDHREVESEKIHMFRRGSEYKH  
NYENNHNHMTPKYFL"

BASE COUNT 692 a 516 c 606 g 672 t

## ORIGIN

Query Match 99.88; Score 2098.2; DB 9; Length 2486;

Best local Similarity 99.98; Pred. No. 0;

Matches 2100; Conservative .0; Mismatches 3; Indels 0; Gaps 0;

1 ATGACCTCGCTGGAGAGAAATCCCTTGGAGTGGCTGCTGGAGTGGTTCTTGCT 60  
183 ATGACCTCGCTGGAGAGAAATCCCTTGGAGTGGCTGCTGGAGTGGTTCTTGCT 242  
61 CTCCTACCATGACCTGGAGCATGATCTATTCTTCTTGGCAACACTAGAACCTCACT 120  
243 CTCCTACCATGACCTGGAGCATGATCTATTCTTCTTGGCAACACTAGAACCTCACT 302  
121 GGGCTTGAAGGTTTGTATAGCATTTCTTCCCAATATCTCTTAACAATTTACTCTTTC 180  
303 GGGCTTGAAGGTTTGTATAGCATTTCTTCCCAATATCTCTTAACAATTTACTCTTTC 362  
181 TGGAAATTTGGTTACAAAGATGATGATACCTGCTGGAGTATTCATTTATGGCAGC 240  
363 TGGAAATTTGGTTACAAAGATGATGATGATACCTGCTGGAGTATTCATTTATGGCAGC 422  
241 AATAGCTCTTCCAGGCTCAATGCAAACTTCAGTATGATGTTCTTCCGCTTGGGCTG 300  
423 AATAGCTCTTCCAGGCTCAATGCAAACTTCAGTATGATGTTCTTCCGCTTGGGCTG 482  
301 TCTTCTCTACTGATGATGCAAGCTGTGACTGTGGTGGGAAATCATTTTGCAGAGTAC 360  
... ..

483 TCTTCTCTACTGATGATGCAAGCTGTGACTGTGGTGGGAAATCATTTTGCAGAGTAC 542  
361 CTCGAAATTTGGGAGATTCATTTTAGACAGATGTTCTTGTGTTCTTACGATATGAT 420  
543 CTCGAAATTTGGGAGATTCATTTTAGACAGATGTTCTTGTGTTCTTACGATATGAT 602  
421 ACTTCACTAATCCCAATCTGAGATTATCAGATGTCACAAAGTATGATGATTAAGT 480  
603 ACTTCACTAATCCCAATCTGAGATTATCAGATGTCACAAAGTATGATGATTAAGT 662  
481 GCCATAGCCACACTTGATGCTATTGGCACAGTGTGACTGACATTAACCTTAAGAAAG 540  
663 GCCATAGCCACACTTGATGCTATTGGCACAGTGTGACTGACATTAACCTTAAGAAAG 722  
541 AAGACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600  
723 AAGACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 782  
601 GCTTTTGGTACCTTGTGTTCTCACCCACCTGCTGTTTGGAGAGTCTCTTGTTC 660  
783 GCTTTTGGTACCTTGTGTTCTCACCCACCTGCTGTTTGGAGAGTCTCTTGTTC 842  
661 AATAGGCGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720  
843 AATAGGCGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 902  
721 GATAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780  
903 GATAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 962  
781 GATAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840  
963 GATAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1022  
841 TGGGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900  
1023 TGGGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1082  
901 ACACCTTGACACCTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 960  
1083 ACACCTTGACACCTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1142  
961 ATATTTTATCTCTGAGAAATATTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020  
1143 ATATTTTATCTCTGAGAAATATTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1202  
1021 GGAGGTGCTACGCTAGAGAAAGATCAGATGCTTTTGGGACATGATGATTAATATC 1080  
1203 GGAGGTGCTACGCTAGAGAAAGATCAGATGCTTTTGGGACATGATGATTAATATC 1262  
1081 GGGCTGAATATGCTATTTTGGCTTGAAGAAACCTTGAATCTCTTCTTCAACAAAGAAC 1140  
1263 GGGCTGAATATGCTATTTTGGCTTGAAGAAACCTTGAATCTCTTCTTCAACAAAGAAC 1322  
1141 AATTCTAAGGCTCTTTCAGAAAGTGAAGAAATACATGAATCTTCTGCTGCTGCT 1200  
1323 AATTCTAAGGCTCTTTCAGAAAGTGAAGAAATACATGAATCTTCTGCTGCTGCT 1382  
1201 GTTGGTGGGAGTTTGGGATTTAGGACATGAGCAATTAAGGAGAAAGTGAAGTGGGC 260  
1383 GTTGGTGGGAGTTTGGGATTTAGGACATGAGCAATTAAGGAGAAAGTGAAGTGGGC 342  
1261 AAGGTGACCAACCAAGAGTCTCTGCTGACATCTGCTTCAAGTTGGATATGAC 1320  
1443 AAGGTGACCAACCAAGAGTCTCTGCTGACATCTGCTTCAAGTTGGATATGAC 1502  
1321 AATGAAGGCTGCTAGCTAGTAAGAAAGTCAAGTCACTGCTCAATGAACAGGTGACAT 1380  
1503 AATGAAGGCTGCTAGCTAGTAAGAAAGTCAAGTCACTGCTCAATGAACAGGTGACAT 1562  
1381 TTTCATTAACATTTTGGAGAGTATGCTTTAAGCCCTATATGGGAAACATGATTAAC 1440  
1563 TTTCATTAACATTTTGGAGAGTATGCTTTAAGCCCTATATGGGAAACATGATTAAC 1622

QY 1441 ATGTGGCTAGGGGAAAAGTTGGGTTCTATACAGACTTGTCCAGACAGAGATTCAC 1500  
 |||||||  
 Db 1623 ATGTGGCTAGGGGAAAAGTTGGGTTCTATACAGACTTGTCCAGACAGAGATTCAC 1682  
 QY 1501 ACTTGGGGATTAATGAGCTTTGTCAAGATACCCAAATTTGTAATCTGAGCATCACTTCCT 1560  
 |||||||  
 Db 1683 ACTTGGGGATTAATGAGCTTTGTCAAGATACCCAAATTTGTAATCTGAGCATCACTTCCT 1742  
 QY 1561 CCGTACCCAGAGGGGAGATTCGACACAGCCATTCACATTCGCTTAACATTTGGGGCAAG 1620  
 |||||||  
 Db 1743 CCGTACCCAGAGGGGAGATTCGACACAGCCATTCACATTCGCTTAACATTTGGGGCAAG 1802  
 QY 1621 CTGTGGATTTTGTCTGACACACTTTGGGAAACAGAGATGACCTCGACAGAAACTG 1680  
 |||||||  
 Db 1803 CTGTGGATTTTGTCTGACACACTTTGGGAAACAGAGATGACCTCGACAGAAACTG 1862  
 QY 1681 CAGGCTATTGCTGTTTCAAAACTACTGAAAAGTACTTAATCAAGTGAATTTTCTGGGA 1740  
 |||||||  
 Db 1863 CAGGCTATTGCTGTTTCAAAACTACTGAAAAGTACTTAATCAAGTGAATTTTCTGGGA 1922  
 QY 1741 TATATCACTTCAGACACTGCTGCTCAGAGATTTCTACAGCTCACTGACATGGCAATGTC 1800  
 |||||||  
 Db 1923 TATATCACTTCAGACACTGCTGCTCAGAGATTTCTACAGCTCACTGACATGGCAATGTC 1982  
 QY 1801 AAGATATGACAGACACTGATCATGACAGATGGTGAATTAATCAATTTGATCGAGGGCTG 1860  
 |||||||  
 Db 1983 AAGATATGACAGACACTGATCATGACAGATGGTGAATTAATCAATTTGATCGAGGGCTG 2042  
 QY 1861 ATCAGGTTGGTTATTCAGAAATCTCCATGCTGTAAGTGAATTCAGAAATTCAGATG 1920  
 |||||||  
 Db 2043 ATCAGGTTGGTTATTCAGAAATCTCCATGCTGTAAGTGAATTCAGAAATTCAGATG 2102  
 QY 1921 GCAAAATTTAGATTCCTCTGATGACCCCACTAATTTATAGAGACACAGAAAGTGTGCTA 1980  
 |||||||  
 Db 2103 GCAAAATTTAGATTCCTCTGATGACCCCACTAATTTATAGAGACACAGAAAGTGTGCTA 2162  
 QY 1981 GACCAAGAGAGATTTCTGAGAAATTTCAATTTATCCGATTTGATGCTCAAAAGAA 2040  
 |||||||  
 Db 2163 GACCAAGAGAGATTTCTGAGAAATTTCAATTTATCCGATTTGATGCTCAAAAGAA 2222  
 QY 2041 GGACACAAATTTATGAAAACAAACATTCATATGATGAAATTCCTCCAAATTCATTTATGA 2100  
 |||||||  
 Db 2223 GGACACAAATTTATGAAAACAAACATTCATATGATGAAATTCCTCCAAATTCATTTATGA 2282  
 C 2101 AAC 2103  
 |||||  
 Db 2283 AAC 2285

RESULT 2  
 LOCUS BC016523 2412 bp mRNA linear ROD 07-AUG-2002  
 DEFINITION Mus musculus, similar to hypothetical protein FLJ21511, clone  
 MGC:27925 IMAGE:3584006, mRNA, complete cds.  
 ACCESSION BC016523  
 VERSION BC016523.1 GI:16741399  
 KEYWORDS MGC.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 Strausberg, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (31-OCT-2001) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA  
 REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 CONTACT Contact: MGC help desk  
 COMMENT Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
 Tissue Procurement: Gilbert Smith, Ph.D.

CDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LILU)  
 DNA Sequencing by: Baylor College of Medicine Human Genome  
 Sequencing Center  
 Center code: BCM-HGSC  
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
 Contact: [amg@bcm.tmc.edu](mailto:amg@bcm.tmc.edu)  
 Gunaratne, P.H., Garcia, A.M., Lu, X., Hu, Y.K., S.W., Hale, S.M.,  
 Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,  
 Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/ILU at: <http://image.llnl.gov>  
 Series: IRAC Plate: 35 Row: k Column: 5  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: Similarity but not  
 identity to protein.

## FEATURES

## source

1..2412  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /map="FVB/N"  
 /clone="MGC:27925 IMAGE:3584006"  
 /tissue\_type="Mammary tumor. Metallothionein-TGF alpha  
 model. 10 month old virgin mouse. Taken by biopsy."  
 /clone\_lib="NCI CGAP\_Mam1"  
 /lab\_host="DH10B"  
 /note="Vector: PCMV-SPORT6"  
 113..2212  
 /codon\_start=1  
 /product="Similar to hypothetical protein FLJ21511"  
 /protein\_id="AAH16523.1"  
 /db\_xref="GI:16741400"

## CDS

/translation="MPGLMRAITLTLILGYVSMYHGLSPMTYFPIQLTELELE  
 FCVAFSLPLTLIPMLLVNKKWTLSLRIYTVGSISFEPANLKLMLVLAGVSS  
 SLIVQYVWMSGGLQRYLKIWGLFGLVHLVLRIMWTSINLPIMVSQMSNVILTL  
 AVAVLDRICTDIDYRNPBGKPREVATGRTSLSSWLLPGAFLSLFLTHMIFGVSI  
 VSRMVGHPHPGDPNPGFGLVLFSSGMLSSSMVLDGALMMMTGAAASAGML  
 YLRMAAVSGCVLAVFTGSMWPVGLVNGSKSGMAMATGCTLLYVLPFCAMCT  
 AFKPYGQYVARESDVILCTIMVITIGLSMLFEPKKNIDFLQTKNSKRTLRCSEKY  
 MKLILMLFVGLGLGRLRHRYEROLGCAATVSAIIMFPRGYONEGPNLERS  
 AQLKEGTADFTLIESDAKRYEINNDLTMVGLKELFVDFGSDRDMGIMVLS  
 RYPIVRSEHNLIPSEGEIAPAITMTVWVSNELVPEVTHFENHDDRLKQIAVLS  
 KLLKCSNOVILFVITSEPSRDYIOLIKHNVDIDISDDRCCEVIMYGLILG  
 VARIHAELSDSEIOMAKFRIPDDPANRDNQKVVIDIRGVKKNIHFNPRCSYKEGH  
 NYENHFRHMTPKIFY"

BASE COUNT 628 a 561 c 616 g 607 t  
 ORIGIN

Query Match 71.8%; Score 1511; DB 10; Length 2412;

Best Local Similarity 82.4%; Pred. 0; Mismatches 370; Indels 0; Gaps 0;

Matches 1733; Conservative 0;

QY 1 ATGACCTGCTGTGGAGAGAAATCTCTTGGAGCTCCGCTGGAGATGTTCTTGTCGTC 60  
 |||||  
 Db 113 ATGCCAGGCTGTGGAGAGAAATCTCTTGGAGCTCCGCTGGAGATGTTCTTGTCGTC 172  
 QY 61 CTTACCATGACCTGGGACCGATGATCTATTTACTTTCCTTTGCAAACTAGACATCACT 120  
 |||||  
 Db 173 CTTACCATGACCTGGGACCGATGATCTATTTACTTTCCTTTGCAAACTAGACATCACT 232  
 QY 121 GGCTGGAAGGTTTGTATGATGATCTCTTCCAAATVTTCTTAACAATTAATCTCTTC 180  
 |||||  
 Db 233 GGCTGGAATTTCTTTGTGTGAGCTTCTCTCCCATCTCTTAACAATTAATCTCTTC 292  
 QY 181 TGGAAATTTGTTAACAAGATGATGATGATGATGATGATGATGATGATGATGATGATG 240  
 |||||  
 Db 293 TGGAAATTTGTTAACAAGATGATGATGATGATGATGATGATGATGATGATGATGATG 352  
 QY 241 ATAGCTCTCTTCAGGCTGCAATGCCAACTGCACTGATGTTCTTGCGCTGGGGGTC 300  
 |||||  
 Db 353 ATAGCTCTCTTCAGGCTGCAATGCCAACTGCACTGATGTTCTTGCGCTGGGGGTC 412



REFERENCE 3 (bases 1 to 184206)  
AUTHORS Waterston, R.H.  
TITLE Direct Submission  
JOURNAL Submitted (05-JAN-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE 4 (bases 1 to 184206)  
AUTHORS Waterston, R.H.  
TITLE Direct Submission  
JOURNAL Submitted (03-JUL-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE 5 (bases 1 to 184206)  
AUTHORS Waterston, R.  
TITLE Direct Submission  
JOURNAL Submitted (07-NOV-2001) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA On Jul 3, 2001 this sequence version replaced g1:12057004.

Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc>  
Contact: [sapiens@wustl.wustl.edu](mailto:sapiens@wustl.wustl.edu)  
Summary Statistics  
Center project name: H\_NH0317G22

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:  
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:  
The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tatenou, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pletzer de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)  
VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:  
The clone sequenced to the left is AC011956. Actual start of this clone is at base position 1 of RP11-317G22; actual end is at base position 184206 of RP11-317G22.

The sequence fidelity of RP11-317G22 between bases 17725 to 17956 can not be guaranteed due to an unresolved dinucleotide repeat. The sequence fidelity of RP11-317G22 between bases 86955 to 87015 can not be guaranteed due to an unresolved homopolymeric run.

FEATURES  
source  
1. 184206  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="4"  
/map="4"  
/clone="RP11-317G22"  
/clone\_1lb="RPCI-11"

repeat\_region 5..70  
/rpt\_family="L2"  
misc\_feature 252..494  
/note="match to EST AA464688 (NID:g2189572) zx82a02.s1"  
misc\_feature 252..374  
/note="match to EST A1077951 (NID:g3412359) cy39a10.s1"  
misc\_feature 252..374  
/note="match to EST A1088656 (NID:g3427715) qb14a11.x1"  
misc\_feature 252..374  
/note="match to EST AW005009 (NID:g5853713) wv02e08.x1"  
misc\_feature 252..374  
/note="match to EST AW006753 (NID:g5855531) wt07a02.x1"  
misc\_feature 252..374  
/note="match to EST BF055447 (NID:g10809343) "  
misc\_feature 252..374  
/note="match to EST BG165990 (NID:g12672693) "  
misc\_feature 252..374  
/note="match to EST BG169511 (NID:g12676214) "  
misc\_feature 252..374  
/note="match to EST BG214746 (NID:g13740767) "  
misc\_feature 252..374  
/note="match to EST BG291592 (NID:g13049843) "  
misc\_feature 252..374  
/note="match to EST W80356 (NID:g1391393) zh49a05.s1"  
misc\_feature 256..374  
/note="similar to Bos taurus EST AV609860 (NID:g9745530) "  
misc\_feature 280..374  
/note="match to EST AW371794 (NID:g6876448) "  
misc\_feature 347..374  
/note="match to EST AA303948 (NID:g1956281) "  
misc\_feature 708..1203  
/note="match to EST B1063872 (NID:g14471399) "  
misc\_feature 1021..1308  
/note="match to EST T06930 (NID:g318079) "  
misc\_feature 1119..1143  
/note="match to EST A1220135 (NID:g3802338) q144g05.x1"  
misc\_feature 1119..1143  
/note="match to EST A1271747 (NID:g3890914) qj77b10.x1"  
misc\_feature 1119..1143  
/note="match to EST BG167015 (NID:g12673728) "  
misc\_feature 1119..1143  
/note="match to EST W68728 (NID:g1377608) zd37c01.s1"  
misc\_feature 1183..1538  
/note="match to EST BG546183 (NID:g13544848) "  
repeat\_region 1319..1350  
/rpt\_family="AT-rich"  
repeat\_region 1349..1466  
/rpt\_family="Alu"  
misc\_feature 1490..1538  
/note="match to EST AA847380 (NID:g2933898) cf07d05.s1"  
misc\_feature 1490..1538  
/note="match to EST A1220135 (NID:g3802338) q144g05.x1"  
misc\_feature 1490..1538  
/note="match to EST A1271747 (NID:g3890914) qj77b10.x1"  
misc\_feature 1490..1538  
/note="match to EST BG167015 (NID:g12673728) "  
misc\_feature 1490..1538  
/note="match to EST W68728 (NID:g1377608) zd37c01.s1"  
misc\_feature 1491..1546  
/note="match to EST BF055447 (NID:g10809343) "  
misc\_feature 1491..1538  
/note="match to EST AA303948 (NID:g1956281) "  
misc\_feature 1491..1538  
/note="match to EST A1077951 (NID:g3412359) cy39a10.s1"  
misc\_feature 1491..1538  
/note="match to EST A1088656 (NID:g3427715) qb14a11.x1"  
misc\_feature 1491..1538  
/note="match to EST AW005009 (NID:g5855713) wv02e08.x1"  
misc\_feature 1491..1538  
/note="match to EST AW006753 (NID:g5855531) wt07a02.x1"  
misc\_feature 1491..1538  
/note="match to EST AW371794 (NID:g6876448) "

```

misc_feature      /note="match to EST BG165990 (NID:g12672693)"
                  1491..1538
                  /note="match to EST BG169511 (NID:g12676214)"
misc_feature      1491..1538
                  /note="match to EST BG214746 (NID:g13740767)"
misc_feature      1491..1538
                  /note="match to EST W80356 (NID:g1391393) zh49a05.s1"
misc_feature      1491..1538
                  /note="match to Bos taurus EST AV609860 (NID:g9745530)"
misc_feature      1491..1538
                  /note="similar to Homo sapiens EST BG207366
                  (NID:g13729053)"
misc_feature      1491..1508
                  /note="match to EST BG291592 (NID:g13049843)"
repeat_region     1561..1606
                  /rpt_family="AT_rich"
repeat_region     1696..1986
                  /rpt_family="Alu"
repeat_region     2175..2217
                  /rpt_family="L1"
repeat_region     2311..2780
                  /rpt_family="MalR"
repeat_region     2828..3176
                  /rpt_family="MalR"
repeat_region     3217..3544
                  /rpt_family="Alu"
repeat_region     3617..3648
                  /rpt_family="AT_rich"
repeat_region     3793..4078
                  /rpt_family="Alu"
repeat_region     4138..4267
                  /rpt_family="Alu"
repeat_region     4462..4655
                  /rpt_family="MERT1-type"
misc_feature      4639..4688

```

Query Match 12.7% Score 267.4; DB 9; Length 184206;  
 Best local Similarity 94.5% Pired. No. 1.8e-63;  
 Matches 277; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

```

QY 799 ACAGAACACCTTCAGCTGGGGGCTCTTACCTGCACACATGGGAGCTGCTGCTCT 858
    |||||||
Db 111216 ACAGAACACCTTCAGCTGGGGGCTCTTACCTGCACACATGGGAGCTGCTGCTCT 111275
    |||||||
QY 859 GGCTGTCTTCCGATCTTACTGATCATGTCGCCCCAACACATGGACACTTATT 918
    |||||||
Db 111276 GGCTGTCTTCCGATCTTACTGATCATGTCGCCCCAACACATGGACACTTATT 111335
    |||||||
QY 919 AACTCAGGACAAACCTGGGAAACCATGACATTCATGATATTTATCTTAGAA 978
    |||||||
Db 111336 AACTCAGGACAAACCTGGGAAACCATGACATTCATGATATTTATCTTAGAA 111395
    |||||||
QY 979 ATATTTTGTGCTGCTGACACGCTTAAAGTTTGTCCAGGAGGTCTACGCTAGA 1038
    |||||||
Db 111396 ATATTTTGTGCTGCTGACACGCTTAAAGTTTGTCCAGGAGGTCTACGCTAGA 111455
    |||||||
QY 1039 GAAAGATCAGATGCTTTGGGACAAATGATTTATTCGGGCTGAATAT 1091
    |||||||
Db 111456 GAAAGATCAGATGCTTTGGGAGTACATTTGAAAGGTCTGATTTAAAT 111508
    |||||||

```

```

RESULT 4
AC108828/c 218040 bp DNA linear HTG 28-APR-2002
LOCUS Mus musculus clone RP23-153H17, WORKING DRAFT SEQUENCE, 21 ordered
DEFINITION
pieces.
ACCESSION AC108828
VERSION AC108828.2 GI:20336119
KEYWORDS HTG: HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Mus musculus.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS

1 (bases 1 to 218040)  
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
 Mus musculus, clone RP23-153H17  
 Unpublished  
 2 (bases 1 to 218040)  
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,  
 Anderson,S., Barina,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,  
 Brown,A., Camarata,J., Campiano,A., Chang,J., Chazaro,B.,  
 Choepel,Y., Colangelo,M., Collins,S., Collumore,A., Cook,A.,  
 Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,  
 Ferreira,P., Fitzhugh,M., Gage,D., Galagan,J., Gardyna,S.,  
 Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
 Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
 Kamat,A., Karatas,A., Kells,C., Labocque,K., Lamazares,R.,  
 Landers,T., Lehoczeky,J., Levine,R., Liu,G., Maclean,C.,  
 Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,  
 McEwen,P., McKernan,K., Melrim,J., Meneus,L., Mihova,T.,  
 Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,  
 Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,  
 Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,  
 Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,  
 Rosetti,M., Roy,A., Santos,R., Schauer,S., Schnupack,R., Seaman,S.,  
 Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
 Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,  
 Topham,K., Travers,M., Travis,N., Triggillo,J., Vassiliev,H.,  
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS

Direct Submission  
 Submitted (31-JAN-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 218040)  
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,  
 Anderson,S., Barina,N., Bastien,V., Bloom,T., Boguslavsky,L.,  
 Boukhgalter,B., Brown,A., Camarata,J., Campiano,A., Chang,J.,  
 Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collumore,A.,  
 Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
 Faro,S., Ferreira,P., Fitzhugh,M., Gage,D., Galagan,J., Gardyna,S.,  
 Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
 Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
 Kamat,A., Karatas,A., Kells,C., Labocque,K., Lamazares,R.,  
 Landers,T., Lehoczeky,J., Levine,R., Lindblad-Toh,K., Liu,G.,  
 Maclean,C., Macdonald,P., Major,J., Margolis,N., Matthews,C.,  
 McCarthy,M., McEwen,P., McKernan,K., Melrim,J., Meneus,L.,  
 Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,  
 Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,  
 Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,  
 Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,  
 Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schnupack,R.,  
 Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
 Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,  
 Topham,K., Travers,M., Travis,N., Triggillo,J., Vassiliev,H.,  
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

COMMENT  
 JOURNAL  
 REFERENCE  
 AUTHORS

Direct Submission  
 Submitted (28-APR-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Apr 28, 2002 this sequence version replaced gi:18450089.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: MIBR  
 Web site: <http://www-seq.wi.mit.edu>  
 Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
 Project Information  
 Center project name: L20971  
 Center clone name: 153\_H17  
 Summary Statistics  
 Sequencing vector: Plasmid; n/a; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960721  
 Consensus quality: 212760 bases at least Q40



Consensus quality: 214579 bases at least Q30  
Consensus quality: 215279 bases at least Q20  
Insert size: 220000; agarose-fp  
Insert size: 216040; sum-of-contigs  
Quality coverage: 7.5 in Q20 bases; agarose-fp  
Quality coverage: 7.6 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 21 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.

This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved.

1  
7064 7163: gap of 100 bp  
7164 8251: contig of 1088 bp in length  
8252 8351: gap of 100 bp  
8352 9118: contig of 767 bp in length  
9119 9218: gap of 100 bp  
9219 10246: contig of 1028 bp in length  
10247 10346: gap of 100 bp  
10347 53500: contig of 43154 bp in length  
53501 53600: gap of 100 bp  
53601 55707: contig of 2107 bp in length  
55708 55807: gap of 100 bp  
55808 55929: contig of 3722 bp in length  
55930 59629: gap of 100 bp  
59630 61934: contig of 2305 bp in length  
61935 62034: gap of 100 bp  
62035 66831: contig of 4797 bp in length  
66832 66931: gap of 100 bp  
66932 72794: contig of 5863 bp in length  
72795 72894: gap of 100 bp  
72895 77248: contig of 4354 bp in length  
77249 77348: gap of 100 bp  
77349 84268: contig of 6920 bp in length  
84269 84368: gap of 100 bp  
84369 93794: contig of 9426 bp in length  
93795 93894: gap of 100 bp  
93895 107682: contig of 1378 bp in length  
107683 107782: gap of 100 bp  
107783 125724: contig of 17942 bp in length  
125725 125824: gap of 100 bp  
125825 137181: contig of 11357 bp in length  
137182 137281: gap of 100 bp  
137282 154808: contig of 17527 bp in length  
154809 154908: gap of 100 bp  
154909 181339: contig of 26431 bp in length  
181340 181439: gap of 100 bp  
181440 195384: contig of 13845 bp in length  
195385 195484: gap of 100 bp  
195485 217207: contig of 21723 bp in length  
217208 217307: gap of 100 bp  
217308 218040: contig of 733 bp in length.

#### FEATURES

1. 218040  
Location/Qualifiers

/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="RP23-153H17"

misc\_feature  
1. 7063  
/note="assembly-fragment"

misc\_feature  
vector\_end:SP6  
vector\_side:left  
7164. 8251

misc\_feature  
/note="assembly-fragment"  
8352. 9118

misc\_feature  
/note="assembly-fragment"  
9219. 10246

misc\_feature  
/note="assembly-fragment"

misc\_feature 10347. 53500  
/note="assembly-fragment"  
misc\_feature 53601. 55707  
/note="assembly-fragment"  
misc\_feature 55808. 59529  
/note="assembly-fragment"  
misc\_feature 59630. 61934  
/note="assembly-fragment"  
misc\_feature 62035. 66831  
/note="assembly-fragment"  
misc\_feature 66932. 72794  
/note="assembly-fragment"  
misc\_feature 72895. 77248  
/note="assembly-fragment"  
misc\_feature 77349. 84268  
/note="assembly-fragment"  
misc\_feature 84369. 93794  
/note="assembly-fragment"  
misc\_feature 93895. 107682  
/note="assembly-fragment"  
misc\_feature 107783. 125724  
/note="assembly-fragment"  
misc\_feature 125825. 137181  
/note="assembly-fragment"  
misc\_feature 137282. 154808  
/note="assembly-fragment"  
misc\_feature 154909. 181339  
/note="assembly-fragment"  
misc\_feature 181440. 195384  
/note="assembly-fragment"  
misc\_feature 195485. 217207  
/note="assembly-fragment"  
misc\_feature 217308. 218040  
/note="assembly-fragment"

BASE COUNT 60131 a 49124 c 47385 g 59388 t 2012 others  
ORIGIN

Query Match 9.2%; Score 192.6; DB 2; Length 218040;  
Best Local Similarity 83.3%; Pred. No. 1.9e-42;

Matches 219; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 799 ACAGAACGCTTCACCTGCGGGGCTCTTACCTGCACACATGGGACGCTGCTGCT 858

Db 67800 ACAGAACGCTTCACCTGCGGGGCTCTTACCTGCACACATGGGACGCTGCTGCT 67741

QY 859 GGCTGTGCTTCGCCATCTTACTGCATGCATGCGCCCAACACCTTGACACTTATT 918

Db 67740 GGATGTGCTCCTGCGCACTTACGGGGTCTATGTGCGCTCAAGTACTTGGCCATCTTGG 67681

QY 919 AACTCAGGCAAAACCTGGGAAAACCATGACCATGATGATTTATCTTCTAGAA 978

Db 67680 AACTCAGGCAAAACCTGGGAAAACCATGACCATGATGATGATTTATCTTCTAGAA 67621

QY 979 AATTTTCTGCTGCTGCGACACGCTTTAAAGTTGTCCAGAGAGTGTCTACGCTAGA 1038

Db 67620 AATTTTCTGCTGCTGCGACACGCTTTAAAGTTGTGTCCAGAGAGTGTCTACGCTAGA 67561

QY 1039 GAAAGATCAGATGCTTTGGG 1061

Db 67560 GAAAGATCAGATGCTTTGGG 67538

RESULT 5  
AC114455/c 162312 bp DNA linear HTG 17-JUL-2002

LOCUS AC114455/c 162312 bp DNA linear HTG 17-JUL-2002

DEFINITION Rattus norvegicus clone CH230-249120, \*\*\* SEQUENCING IN PROGRESS

ACCESSION AC114455  
VERSION AC114455.6 GI:21745639

KEYWORDS HTG; HTGS; PHASE1.  
SOURCE Norway rat.

## ORGANISM

Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

REFERENCE  
AUTHORS

1 (bases 1 to 162312)  
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,  
Alsprouks,S.L., Amaralunge,H.C., Are,J.R., Ayale,M., Banks,T.,  
Barbata,J., Benton,J., Bimege,K., Blankenburg,K., Bonnin,D.,  
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,  
Carron,I.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,  
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,  
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,  
DeVella,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,  
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,  
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,  
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,  
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,  
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,  
Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K.,  
Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J.,  
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,  
Homel,F., Howard,S., Huber,J., Huliyk,S., Hume,J., Jackson,L.E.,  
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,  
Karlsom,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,  
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,  
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louised,H.,  
Lozdo,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,  
Maheshwari,M., Kapua,P., Martin,R., Martindale,A., Martinez,E.,  
Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mel,G., Metzger,M.,  
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,  
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,  
Nguyen,N., Nickerson,E., Nwokenwo,S., Ogih,M., Okunou,G.,  
Oragunye,N., Oviedo,R., Pace,A., Payton,P., Peery,J., Perez,L.,  
Peterson,L., Pickens,R., Primus,E., Pul,L.L., Quiles,M., Ren,Y.,  
Rivers,M., Rojas,A., Rojudoan,I., Rolfe,M., Ruiz,S., Saverly,G.,  
Scherrer,S., Scott,G., Shen,H., Shoochari,N., Sisson,I.,  
Sodergren,E., Sonaik,T., Sparks,A., Stanley,H., Stone,H.,  
Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K., Tang,H.,  
Taney,J., Taylor,C., Taylor,T., Telford,B., Thomas,R., Thomas,S.,  
Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,  
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S.,  
Williams,G., Williamson,A., Wleczky,R., Wooden,S., Wolley,K.,  
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,  
Weinstock,G. and Gibbs,R.

TITLE  
JOURNALREFERENCE  
AUTHORSTITLE  
JOURNALREFERENCE  
AUTHORSTITLE  
JOURNAL

## COMMENT

2 (bases 1 to 162312)  
Worley,K.C.  
Direct Submission  
Submitted (09-MAR-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 162312)  
Worley,K.C.  
Direct Submission  
Submitted (17-JUL-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Jul 14, 2002 this sequence version replaced gi:20467645.  
----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GSX1  
Center clone name: CH230-249120  
----- Summary Statistics  
Sequencing vector: Plasmid:  
Chemistry: Dye-terminator Big Dye: 100% of reads  
Assembly program: Phrap: version 0.990329  
Consensus quality: 12635 bases at least Q40

Consensus quality: 126261 bases at least Q30  
Consensus quality: 129845 bases at least Q20

NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
NOTE: This is a "working draft" sequence. It currently  
consists of 49 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.

1	1174:	contig of 1174 bp in length
1175	1274:	gap of unknown length
1275	2899:	contig of 1625 bp in length
2900	2999:	gap of unknown length
3000	4263:	contig of 1264 bp in length
4264	4363:	gap of unknown length
4364	5667:	contig of 1304 bp in length
5668	5767:	gap of unknown length
5768	7230:	contig of 1463 bp in length
7231	7330:	gap of unknown length
7331	8913:	contig of 1583 bp in length
8914	9013:	gap of unknown length
9014	10114:	contig of 1101 bp in length
10115	10214:	gap of unknown length
10215	11240:	contig of 1026 bp in length
11241	11340:	gap of unknown length
11341	12994:	contig of 1654 bp in length
12995	13094:	gap of unknown length
13095	14158:	contig of 1064 bp in length
14159	14258:	gap of unknown length
14259	15405:	contig of 1147 bp in length
15406	15505:	gap of unknown length
15506	17366:	contig of 1861 bp in length
17367	17466:	gap of unknown length
17467	18850:	contig of 1384 bp in length
18851	18950:	gap of unknown length
18951	21635:	contig of 2655 bp in length
21636	21735:	gap of unknown length
21736	22866:	contig of 1131 bp in length
22867	22966:	gap of unknown length
22967	24541:	contig of 1575 bp in length
24542	24641:	gap of unknown length
24642	26761:	contig of 2120 bp in length
26762	26861:	gap of unknown length
26862	28689:	contig of 1828 bp in length
28690	28789:	gap of unknown length
28790	31195:	contig of 2406 bp in length
31196	31295:	gap of unknown length
31296	34187:	contig of 2892 bp in length
34188	34287:	gap of unknown length
34288	37140:	contig of 2853 bp in length
37141	37240:	gap of unknown length
37241	38773:	contig of 1533 bp in length
38774	38873:	gap of unknown length
38874	41274:	contig of 2401 bp in length
41275	41374:	gap of unknown length
41375	42704:	contig of 1330 bp in length
42705	42804:	gap of unknown length
42805	44612:	contig of 1808 bp in length
44613	44712:	gap of unknown length
44713	47204:	contig of 2492 bp in length
47205	47304:	gap of unknown length
47305	49564:	contig of 2260 bp in length
49565	49664:	gap of unknown length
49665	52047:	contig of 2383 bp in length
52048	52147:	gap of unknown length
52148	55151:	contig of 3004 bp in length
55152	55251:	gap of unknown length
55252	58346:	contig of 3095 bp in length
58347	58446:	gap of unknown length

*	58487	61931: contig of 3485 bp in length
*	61932	62018: gap of unknown length
*	62032	65703: contig of 3687 bp in length
*	65719	65818: gap of unknown length
*	65819	65981: contig of 4173 bp in length
*	65992	70091: gap of unknown length
*	70092	73876: contig of 3785 bp in length
*	73877	73976: gap of unknown length
*	73977	77758: contig of 3782 bp in length
*	77759	77858: gap of unknown length
*	77859	80356: contig of 2498 bp in length
*	80357	80456: gap of unknown length
*	80457	86083: contig of 5627 bp in length
*	86084	86183: gap of unknown length
*	86184	90591: contig of 4408 bp in length
*	90592	90691: gap of unknown length
*	90692	94431: contig of 3730 bp in length
*	94422	94531: gap of unknown length
*	94522	99012: contig of 4491 bp in length
*	99013	99112: gap of unknown length
*	99113	102900: contig of 3788 bp in length
*	102901	103000: gap of unknown length
*	103001	106277: contig of 3277 bp in length
*	106278	106377: gap of unknown length
*	106378	111333: contig of 4976 bp in length
*	111354	111453: gap of unknown length
*	111454	118794: contig of 7341 bp in length
*	118795	118894: gap of unknown length
*	118895	124281: contig of 5287 bp in length
*	118895	124281: gap of unknown length
*	124281	133604: contig of 9323 bp in length
*	133605	133704: gap of unknown length
*	133705	141028: contig of 7324 bp in length
*	141029	141128: gap of unknown length
*	141129	149603: contig of 8475 bp in length
*	149604	149703: gap of unknown length
*	149704	162312: contig of 12609 bp in length.

BASE COUNT	40948 a	34779 c	33666 g	42210 t	10709 others
ORIGIN					
Query Match	9.1%	Score 191:	DB 2:	Length 162312;	
Local Similarity	82.9%;	Pred. No. 5.3e-42;			
Matches 218; Conservative	0;	Mismatches 45;	Indels 0;	Gaps 0;	

QY	799	ACGGACAGCTGCTACGTGCGGGGCGCTCTTTACCTGACACATCGGAGACCTGCTGTCT	858
Db	59064	ACAGSAGCACTTCACGCCATGGGTCTCCTTTATCTGCGCAGCTTTGGGACGCCGCTGTTCT	59005
QY	859	GCGTGTGTCCTTGCCCATCTTTACTGCATTCATGTGAGCCCAACACTTGGACACCTTATT	918
Db	59004	GGATCGCAGCTGGCGAGTCTTCACAGGGTCCATGTGGCTCAGTACTTGGACATCATTTAG	58945
QY	919	AACATGAGGACAAACCCCTGGGAAACCATGACCATTTGCCATATATTTTATCTTTAGAA	978
Db	58944	AACATGAGGACAAACCCCTGGAGACCCATGACACATCGCATGATGAGCTTATTTCTTACAA	58885
QY	979	ATATTTTCTGTGCGCTGGTGACACACTTTTAGTTTGCCAGAGAGTGTCTACGCTAGA	1038
Db	58884	GTGTTTCTGTGCGCTGGTGACACACTTTTAGTTTGTTCTCGAGAGGTGTCTACGCTAGA	58825
QY	1039	GAAAGATCAGATGTGCTTTTGGG	1061
Db	58824	GAAAGTCTGATGTGCTTTTGGG	58802

DEFINITION	homo sapiens chromosome 6 clone RP3-405B4, *** SEQUENCING IN PROGRESS ***, 33 unordered pieces.
ACCESSION	AL158816
VERSION	AL158816.11 GI:9943390
KEYWORDS	HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE	human.
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	1 (bases 1 to 222885)
AUTHORS	Sims, S.
TITLE	Direct Submission
JOURNAL	Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
COMMENT	requests: clonerequest@sanger.ac.uk On Aug 29, 2000 this sequence version replaced gi:9926474.

```
* 87340 93671: contig of 6332 bp in length
* 93672 93771: gap of 100 bp
* 93772 97550: contig of 3779 bp in length
* 97551 97650: gap of 100 bp
* 97651 100101: contig of 2451 bp in length
* 100102 100201: gap of 100 bp
* 100202 103416: contig of 3215 bp in length
* 103417 103516: gap of 100 bp
* 103517 107564: contig of 4048 bp in length
* 107565 107664: gap of 100 bp
* 107665 131778: contig of 24114 bp in length
* 131779 131878: gap of 100 bp
* 131879 135261: contig of 3383 bp in length
* 135262 135361: gap of 100 bp
* 135362 137913: contig of 2552 bp in length
* 137914 138013: gap of 100 bp
* 138014 140949: contig of 2936 bp in length
* 140950 141049: gap of 100 bp
* 141050 156466: contig of 15417 bp in length
* 156467 156566: gap of 100 bp
* 156567 164836: contig of 8270 bp in length
* 164837 164936: gap of 100 bp
* 164937 168950: contig of 4014 bp in length
* 168951 169050: gap of 100 bp
* 169051 173724: contig of 4674 bp in length
* 173725 173824: gap of 100 bp
* 173825 179991: contig of 6167 bp in length
* 179992 180091: gap of 100 bp
* 180092 190208: contig of 10117 bp in length
* 190209 190308: gap of 100 bp
* 190309 193471: contig of 3163 bp in length
* 193472 193571: gap of 100 bp
* 193572 197829: contig of 4258 bp in length
* 197830 197929: gap of 100 bp
* 197930 222885: contig of 24956 bp in length.
```

## FEATURES

## SOURCE

```
1. 222885
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="RP3-405B4"
/clone_id="RPC1-3"
1. 3599
/note="assembly-fragment:00027"
misc_feature
3700..12818
/note="assembly-fragment:00056"
misc_feature
12919..16398
/note="assembly-fragment:00093"
misc_feature
16499..18590
/note="assembly-fragment:00125"
misc_feature
18691..22372
/note="assembly-fragment:00140"
misc_feature
22473..26905
/note="assembly-fragment:00167"
misc_feature
27006..29581
/note="assembly-fragment:00175"
misc_feature
29682..33446
/note="assembly-fragment:00178"
misc_feature
33547..48536
/note="assembly-fragment:00416"
misc_feature
48637..55599
/note="assembly-fragment:00542"
misc_feature
55700..65124
/note="assembly-fragment:00565"
misc_feature
65225..69090
/note="assembly-fragment:00583"
misc_feature
69191..73894
/note="assembly-fragment:00649"
misc_feature
73995..82665
/note="assembly-fragment:00709"
misc_feature
82766..87239
/note="assembly-fragment:00789"
misc_feature
87340..93671
```

```
/note="assembly-fragment:00874"
93772..97550
/note="assembly-fragment:00924"
97651..100101
/note="assembly-fragment:01052"
100202..103416
/note="assembly-fragment:01082"
103517..107564
/note="assembly-fragment:01216"
107665..131778
/note="assembly-fragment:01234"
131879..135261
/note="assembly-fragment:01280"
135362..137913
/note="assembly-fragment:01337"
138014..140949
/note="assembly-fragment:01643"
141050..156466
/note="assembly-fragment:01660"
156567..164836
/note="assembly-fragment:01690"
164937..168950
/note="assembly-fragment:01692.0"
169051..173724
/note="assembly-fragment:01748"
173825..179991
/note="assembly-fragment:02061"
180092..190208
/note="assembly-fragment:02140"
190309..193471
/note="assembly-fragment:02150"
193572..197829
/note="assembly-fragment:02181"
197930..222885
/note="assembly-fragment:00834
clone_end:T7
vector_side:right"
```

```
BASE COUNT 69666 a 41364 c 44947 g 63682 t 3226 others
```

```
Query Match 7.6% Score 160, DB 2; Length 222885;
Best Local Similarity 89.6%; Pred. No. 2,8e-33;
Matches 172; Conservative 0; Mismatches 20; Indels 0; Gaps 0;
```

```
QY 1863 CAGGTGGTTATGCAAGATTCCTCCATGCTGAAGTACGATTCGAATTCAGATGCG 1922
|||||
DB 157674 CAGGTGGTTATGCAAGATTCCTCCATGCTGAAGTACGATTCGAATTCAGATGCG 157733
|||||
QY 1923 AAATTAGCATCCCTGATGACCCCACTAATTATAGAGACACAGAAAGTGCATAGA 1982
|||||
DB 157734 AAATTAGCATCCCTGATGACCCCACTAATTATAGAGACACAGAAAGTGCATAGA 157793
|||||
QY 1983 CCACAGAGAGTTTCTGAGAAAATTCATTTTAATCCAGATTTGATCCTACAAAGAAG 2042
|||||
DB 157794 CCACAGAGAGTTTCTGAGAAAATTCATTTTAATCCAGAGTGCCTTATGCTGTAG 157853
|||||
QY 2043 ACACAATTATGA 2054
|||||
DB 157854 TTCAGATTATGA 157865
|||||
```

```
RESULT 7
AL158816/c 222885 bp DNA linear HTG 10-JUL-2001
LOCUS AL158816 Homo sapiens chromosome 6 clone RP3-405B4, *** SEQUENCING IN
DEFINITION PROGRESS ***, 33 unordered pieces.
ACCESSION AL158816
VERSION AL158816.11 GI:9943990
KEYWORDS HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.*
```

REFERENCE 1 (bases 1 to 222885)  
 AUTHORS Sims, S.  
 TITLE Direct Submission  
 JOURNAL Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
 requests: clonerequest@sanger.ac.uk  
 On Aug 29, 2000 this sequence version replaced gi:9926474.  
 ----- Genome Center  
 Center: Sanger Centre  
 Center code: SC  
 Web site: http://www.sanger.ac.uk  
 Contact: humquerry@sanger.ac.uk  
 ----- Project Information  
 Center project name: dj405B4  
 ----- Summary Statistics  
 Assembly program: XGAP4: version 4.5  
 Sequencing vector: plasmid, L08752; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Consensus quality: 204390 bases at least Q40  
 Consensus quality: 211527 bases at least Q30  
 Consensus quality: 215491 bases at least Q20  
 Insert size: 219685; sum-of-contigs  
 Insert size: 126549; 14.0% error; agarose-fp  
 Quality coverage: 3.63x in Q20 bases; sum-of-contigs Quality  
 coverage: 6.72x in Q20 bases; agarose-fp  
 -----

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 33 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

\* 1  
 \* 3600 3699: contig of 3599 bp in length  
 \* 3700 12818: contig of 9119 bp in length  
 \* 12819 12918: gap of 100 bp  
 \* 12919 16398: contig of 3480 bp in length  
 \* 16399 16498: gap of 100 bp  
 \* 16499 18590: contig of 2092 bp in length  
 \* 18591 18690: gap of 100 bp  
 \* 18691 22372: contig of 3682 bp in length  
 \* 22373 22472: gap of 100 bp  
 \* 22473 26905: contig of 4433 bp in length  
 \* 26906 27005: gap of 100 bp  
 \* 27006 29581: contig of 2576 bp in length  
 \* 29582 29681: gap of 100 bp  
 \* 29682 33446: contig of 3765 bp in length  
 \* 33447 33546: gap of 100 bp  
 \* 33547 48536: contig of 14990 bp in length  
 \* 48537 48636: gap of 100 bp  
 \* 48637 55599: contig of 6963 bp in length  
 \* 55600 55699: gap of 100 bp  
 \* 55700 65124: contig of 9425 bp in length  
 \* 65125 65224: gap of 100 bp  
 \* 65225 69090: contig of 3866 bp in length  
 \* 69091 69190: gap of 100 bp  
 \* 69191 73894: contig of 4704 bp in length  
 \* 73895 73994: gap of 100 bp  
 \* 73995 82665: contig of 8671 bp in length  
 \* 82666 82765: gap of 100 bp  
 \* 82766 87239: contig of 4474 bp in length  
 \* 87240 87339: gap of 100 bp  
 \* 87340 93671: contig of 6332 bp in length  
 \* 93672 93771: gap of 100 bp  
 \* 93772 97550: contig of 3779 bp in length  
 \* 97551 97650: gap of 100 bp  
 \* 97651 100101: contig of 2451 bp in length  
 \* 100102 100201: gap of 100 bp  
 \* 100202 103416: contig of 3215 bp in length  
 \* 103417 103516: gap of 100 bp  
 \* 103517 107564: contig of 4048 bp in length

## FEATURES

## source

\* 107565 107664: gap of 100 bp  
 \* 107665 131778: contig of 24114 bp in length  
 \* 131779 131878: gap of 100 bp  
 \* 131879 135261: contig of 3383 bp in length  
 \* 135262 135361: gap of 100 bp  
 \* 135362 137913: contig of 2552 bp in length  
 \* 137914 138013: gap of 100 bp  
 \* 138014 140949: contig of 2936 bp in length  
 \* 140950 141049: gap of 100 bp  
 \* 141050 156466: contig of 15417 bp in length  
 \* 156467 156566: gap of 100 bp  
 \* 156567 164836: contig of 8270 bp in length  
 \* 164837 164936: gap of 100 bp  
 \* 164937 168950: contig of 4014 bp in length  
 \* 168951 169050: gap of 100 bp  
 \* 169051 173724: contig of 4674 bp in length  
 \* 173725 173824: gap of 100 bp  
 \* 173825 179991: contig of 6167 bp in length  
 \* 179992 180091: gap of 100 bp  
 \* 180092 190208: contig of 10117 bp in length  
 \* 190209 190308: gap of 100 bp  
 \* 190309 193471: contig of 3163 bp in length  
 \* 193472 193571: gap of 100 bp  
 \* 193572 197829: contig of 4238 bp in length  
 \* 197830 197929: gap of 100 bp  
 \* 197930 222885: contig of 24956 bp in length.  
 -----  
 location/Qualifiers  
 1. 222885  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="6"  
 /clone\_lib="RP3-405B4"  
 /clone\_lib="RPCI-3"  
 1. 3539  
 /note="assembly-fragment:00027"  
 3700. 12818  
 /note="assembly-fragment:00056"  
 12919. 16398  
 /note="assembly-fragment:00093"  
 16499. 18590  
 /note="assembly-fragment:00125"  
 18691. 22372  
 /note="assembly-fragment:00140"  
 22473. 26905  
 /note="assembly-fragment:00167"  
 27006. 29581  
 /note="assembly-fragment:00175"  
 29682. 33446  
 /note="assembly-fragment:00178"  
 33547. 48536  
 /note="assembly-fragment:00416"  
 48637. 55599  
 /note="assembly-fragment:00542"  
 55700. 65124  
 /note="assembly-fragment:00565"  
 65225. 69090  
 /note="assembly-fragment:00583"  
 69191. 73894  
 /note="assembly-fragment:00649"  
 73995. 82665  
 /note="assembly-fragment:00709"  
 82766. 87239  
 /note="assembly-fragment:00789"  
 87340. 93671  
 /note="assembly-fragment:00874"  
 93772. 97550  
 /note="assembly-fragment:00924"  
 97651. 100101  
 /note="assembly-fragment:01052"  
 100202. 103416  
 /note="assembly-fragment:01082"  
 103517. 107564  
 /note="assembly-fragment:01216"

```

misc_feature      107665..131778
                  /note="assembly_fragment:01234"
misc_feature      131879..135261
                  /note="assembly_fragment:01280"
misc_feature      135362..137913
                  /note="assembly_fragment:01337"
misc_feature      138014..140949
                  /note="assembly_fragment:01643"
misc_feature      141050..156466
                  /note="assembly_fragment:01660"
misc_feature      156567..164836
                  /note="assembly_fragment:01690"
misc_feature      164937..168950
                  /note="assembly_fragment:01692.0"
misc_feature      169051..173724
                  /note="assembly_fragment:01748"
misc_feature      173825..179991
                  /note="assembly_fragment:02061"
misc_feature      180092..190208
                  /note="assembly_fragment:02140"
misc_feature      190309..193471
                  /note="assembly_fragment:02150"
misc_feature      193572..197829
                  /note="assembly_fragment:02181"
misc_feature      197930..222885
                  /note="assembly_fragment:00834
                  vector_side:right"
BASE COUNT      69666 a 41364 c 44947 g 63682 t 3226 others
ORIGIN
Query Match
Best Local Similarity 100.0%; Pred. No. 9,4e-31;
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1508 GGATTATGCGTTTCTCAAGATACCAATTGTAATCTGAGCATCACCCTTCTCCGTAC 1567
      |||||||
Db 215729 GGATTATGCGTTTCTCAAGATACCAATTGTAATCTGAGCATCACCCTTCTCCGTAC 215670
QY 1568 CAGAGGCGGAGATCGCACCCATCATCATTTGACCGTTTACATTTGGGGCAAGCTGTGG 1627
      |||||||
Db 215669 CAGAGGCGGAGATCGCACCCATCATCATTTGACCGTTTACATTTGGGGCAAGCTGTGG 215610
QY 1628 ATTTTGTGTCGACACACTTTGGGAACACGA 1658
      |||||||
Db 215609 ATTTTGTGTCGACACACTTTGGGAACACGA 215579

```

```

PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
1580102
3 (bases 1 to 316613)
Wilson,C., Bergantino,E., Lanfranchi,G., Valle,G., Carignani,G. and Frontali,L.
A putative serine/threonine protein kinase gene on chromosome III of Saccharomyces cerevisiae
Yeast 8 (1), 71-77 (1992)
92254506
1580103
4 (bases 1 to 316613)
Bentl,P., Chanet,R., Fabre,F., Faye,G., Fukuhara,H. and Sor,F.
Sequence of the sup61-RAD18 region on chromosome III of Saccharomyces cerevisiae
Yeast 8 (2), 147-153 (1992)
92221691
JOURNAL
MEDLINE
PUBMED
1561837
5 (bases 1 to 316613)
Bolle,P.A., Gilliquet,V., Berben,G., Dumont,J. and Hlilger,F.
The complete sequence of K3B, a 7.9 kb fragment between PKI and CRY1 on chromosome III, reveals the presence of seven open reading frames
Yeast 8 (3), 205-213 (1992)
92245758
JOURNAL
MEDLINE
PUBMED
1574926
6 (bases 1 to 316613)
Sor,F., Cheret,G., Fabre,F., Faye,G. and Fukuhara,H.
Sequence of the HMR region on chromosome III of Saccharomyces cerevisiae
Yeast 8 (3), 215-222 (1992)
92245759
JOURNAL
MEDLINE
PUBMED
1574927
7 (bases 1 to 316613)
Oliver,S.G., van der Aart,Q.J., Agostoni-Carbone,M.L., Aigle,M., Alberghina,L., Alexandrakl,D., Antoine,G., Anwar,R., Ballesta,J.P., Bentl,P. et al.
The complete DNA sequence of yeast chromosome III
Nature 357 (6373), 38-46 (1992)
92244356
JOURNAL
MEDLINE
PUBMED
1574125
8 (bases 1 to 316613)
Skala,J., Purnelle,B. and Goffeau,A.
The complete sequence of a 10.8 kb segment distal of SUP2 on the right arm of chromosome III from Saccharomyces cerevisiae reveals seven open reading frames including the RVS161, ADP1 and PGK genes
Yeast 8 (5), 409-417 (1992)
92327849
JOURNAL
MEDLINE
PUBMED
1626432
9 (bases 1 to 316613)
Wilson,C., Grisanti,P. and Frontali,L.
The complete sequence of a 6146 bp fragment of Saccharomyces cerevisiae chromosome III contains two new open reading frames
Yeast 8 (7), 569-575 (1992)
92397594
JOURNAL
MEDLINE
PUBMED
1523889
10 (bases 1 to 316613)
Scherens,B., Messenguy,F., Gigot,D. and Dubois,E.
The complete sequence of a 9,543 bp segment on the left arm of chromosome III reveals five open reading frames including glucokinase and the protein disulfide isomerase
Yeast 8 (7), 577-585 (1992)
92397595
JOURNAL
MEDLINE
PUBMED
1523890
11 (bases 26740 to 32076)
Defoor,E., Debrabandere,R., Keyers,B., Voel,M. and Volckaert,G.
Nucleotide sequence of D10B, a BamHI fragment on the small-ring chromosome III of Saccharomyces cerevisiae
Yeast 8 (8), 681-687 (1992)
93070606
JOURNAL
MEDLINE
PUBMED
1441748
12 (bases 169581 to 171116; 171683 to 172169)
Agostoni Carbone,M.L., Panzerl,L., Muzi Falconi,M., Carcano,C., Plevani,P. and Lucchini,G.
Nucleotide sequence of 9.2 kb left of CRY1 on yeast chromosome III

```











```
misc_feature
/note="ttaacataag, splice branch and acceptor"
/complement(2103..2108)
/gene="SPAC688.02c"
/misc_feature
/note="gtgagt, splice donor sequence"
/complement(2257..2271)
/gene="SPAC688.02c"
/misc_feature
/note="ttaacatttaag, splice branch and acceptor"
/complement(2337..2342)
/gene="SPAC688.02c"
/misc_feature
/note="gtgtgt, splice donor sequence"
/complement(join(2596..2953,3083..3297,3381..3422))
/gene="SPAC688.03c"
/complement(join(2596..2953,3083..3297,3381..3422))
/gene="SPAC688.03c"
/note="SPAC688.03c, len:204, SIMILARITY: Homo sapiens,
O914X0, amecr1 protein, (333 aa), fasta scores: opt:
443, E():3.4e-25, (40.3% identity in 186 aa)"
/codon_start=1
/label="SPAC688.03c
/product="conserved hypothetical protein"
/protein_id="CAB90770.1"
/db_xref="GI:7768479"
/translation="MICWPSLNKESWMDKEYCYCEVVAATLEHRKVRDKNWAKSWT
RSIPLVKFAFGKHDKQLRGICIGFRAPLVNTLYFSKQAFCDERFRISIGELA
LLECOIDLVDFEIDPLDWEVGHGVSIFKFTANGIYSTYLPVSAEQRMDOEET
LESLHKAGYGSIRSLQITATPKYSLEIGCTYEELHNLELIG"
/complement(join(2662..2953,3083..3273))
/gene="SPAC688.03c"
/misc_feature
/note="Match to PF01871 DUF51, Protein of unknown function
Score 49.41"
/complement(2954..2976)
/gene="SPAC688.03c"
/note="ctaatacaacatttaaaataag, splice branch and
acceptor"
/complement(3077..3082)
/gene="SPAC688.03c"
/note="gtaaagt, splice donor sequence"
/complement(3298..3315)
/gene="SPAC688.03c"
/misc_feature
/note="ctaatacaatggacaag, splice branch and acceptor"
/complement(3375..3380)
/gene="SPAC688.03c"
/misc_feature
/note="gtatgga, splice donor sequence"
/complement(4047..4775)
/gene="SPAC688.04c"
/complement(4047..4775)
/gene="SPAC688.04c"
/note="SPAC688.04c, len:242, SIMILARITY: Saccharomyces
cerevisiae, Y18_YEAST, Glutathione transferase GT11, (234
aa), fasta scores: opt: 274, E():3.5e-11, (27.0% identity
in 222 aa)"
/codon_start=1
/label="SPAC688.04c
/product="putative glutathione S transferase"
/protein_id="CAB90771.1"
/db_xref="GI:7768480"
/translation="MIVLHLKNSRSTIYMLLETKPYEIKYDRVDRAPRAYTK
LSPIGKSTIVDDGVYTESAILEHLVKKGPSRSEEDVALEKLELMHSEAS
LMPFTWASHVLDLSNMPIFERIYVROFVNGIKSKYLSKETFLVDLIDNHLASNEA
FAGQDFTAADPQMFPIFAQRDYLSQPKYNIKRMVSDRPACRIAAKEVDENLT
ILFSDVERYSHPPTPEQVRSD"
/complement(4164..4775)
/gene="SPAC688.04c"
/misc_feature
/note="Match to PF00043 GST, Glutathione S-transferases.
Score 85.13"
/complement(join(5637..6842,6927..6980))
/gene="SPAC688.06c"
/complement(join(5637..6842,6927..6980))
/gene="SPAC688.06c"
/note="SPAC688.06c, len:419, predicted to contain
coiled-coil region, SIMILARITY: S. cerevisiae, Low to
Actin-related protein, regulator of microtubule stability,
YB61_YEAST, (324 aa), fasta scores: opt: 125, E():0.76,
```

```
(21.3% identity in 300 aa)"
/codon_start=1
/label="SPAC688.06c
/product="hypothetical protein very putative Actin-related
protein, regulator of microtubule stability"
/protein_id="CAB90772.1"
/db_xref="GI:7768481"
/translation="MSAEYIEVSSSPDIFDDDMITFPELNKPKDKSKRRSY
TECCERILITSKDPESTFOQLVHNHCTSHKHENHLNVADEDEDETLPLFSSFD
NESDLEPDLNTRVAEDNDVLISRYSKTNSCRNFEHSAHSNREISSGYYH
KKPOLFEKSLKLGKNSIEANRSPILKEICSANSTENVCFSVDEIQORHSAGH
SIDTQONSNEFLESDATKRKRTDNKEFTSCFNDRSRTLNTAGYMDINKADNE
AKSLKEKLENPVEKRLAIASYGKSSDSKATLIKYESCDAIDSOSKTKGKET
PHDYLITSTKTVLFRDDITVQTHRAISQVVOADNDSWIMILITYSIDVBEFOLMLK
RKNLNVSLDLKSMCDKIVGLMKSMH"
/complement(6843..6857)
/gene="SPAC688.06c"
/misc_feature
/note="ttaactgagcttag, splice branch and acceptor"
/complement(6921..6926)
/gene="SPAC688.06c"
/note="gtatgga, splice donor sequence"
/complement(7359..10475)
gene
Query Match 5.8%; Score 121.8; DB 8; Length 34034;
Best Local Similarity 49.0%; Pred. No. 1.3e-22;
Matches 324; Conservative 0; Mismatches 337; Indels 0; Gaps 0;
misc_feature
1283 TCTCTGCTGCATCTGGCCTTTCAGGTTTGATATGACATGAAGGCTGCTAGTCTAG 1342
11 111111 111111 111111 111111 111111
481 TCACCTGCTGCTATTTGAGATATTCATCTTGGCTAGACATTTCAATGATGACAGAGA 540
11 111111 111111 111111 111111 111111
1343 AAGATCAGCTCAGCCGCTCATGAAACAGGTCGAGATTTCATCAATTTGGACAGAG 1402
11 111111 111111 111111 111111 111111
541 ATGCTATACGCATGCTGTCGAGATATGAACTGATCTCTTGGCTACTAGGATCTG 600
11 111111 111111 111111 111111 111111
1403 ATGCTTCTAGGCTATATATGGGGAACAATGACTTACATCTGGCTGAGGGAAGATTGG 1462
11 111111 111111 111111 111111 111111
601 ACACACACAGTTTAATTTATGCGATTTTCGAGATTTAAACAGGTTTACGACATGATCTAG 660
11 111111 111111 111111 111111 111111
1463 GTTCTATACAGACTTTGGTCCAGACCAAGGTATCACACTTGGGGGATTTATGCTTTGT 1522
11 111111 111111 111111 111111 111111
661 GAATGATGCGGATTTATGGCCAGCCGATTAACATCACTACGCGGCTGCTTACTCT 720
11 111111 111111 111111 111111 111111
1523 CAAGATACCCATTTGGAATCTGACATCACCCTTCCGTCACCAAGGCGGAGATCG 1582
11 111111 111111 111111 111111 111111
721 CCAAAATTTCTATTTGCAACTCGACTCACCATTTGTTACCTTCTCTCAAGAGAACTTG 780
11 111111 111111 111111 111111 111111
1583 CACACGCCATCATTCATGACCGTTTAACATTTCCGGGCAAGCTGGGATTTGTCGTACAC 1642
11 111111 111111 111111 111111 111111
781 CTCCTGCTATTTGATGACACATGATGCTACGAGAGCAATTAAGATGTTGCTTCTC 840
11 111111 111111 111111 111111 111111
1643 ACTTTGGGACCAAGATGACCTCGACAGGAACATCGAGGCTATTTGCTTTCAAAAC 1702
11 111111 111111 111111 111111 111111
841 ACAATGCTCAATATGAACCTCAGTTGATGACGCTTCAAGACATGAGTGGCTCGAA 900
11 111111 111111 111111 111111 111111
1703 TACTGAAAGTGCCTATATCAAGTATATTTCTGGGATATATCACTTACGACCTGGCT 1762
11 111111 111111 111111 111111 111111
901 TTATGCGCGAAGATCCACAGCCCTTATGCTTCTCGGAATATGTTGTTAAACGTTGGCC 960
11 111111 111111 111111 111111 111111
1763 CCAGAGATTATGTACAGCTCCTACATGGAACATGGAATTTGAAGATATCGACAGCTGATC 1822
11 111111 111111 111111 111111 111111
961 AAGAACCACAATATTATTAACCGGACACGGGATATTGAACATTTGAACCGCTGACT 1020
11 111111 111111 111111 111111 111111
1823 ATGACAGATGCTGTGAATACATTAATGATATGACAGGCTGATCAGCTGGTATGCAAGA 1882
11 111111 111111 111111 111111 111111
1021 ATGATGCTGTGTGATGATATATTTATTTATGAGGTGTCAAGCAATTTGGTATGCTGAG 1080
11 111111 111111 111111 111111 111111
1883 TGTCCATGCTGATCAGATGATGATGAAATTCAGATGCGCAAAATTTAGATCCCTGATG 1942
11 111111 111111 111111 111111 111111
1081 TTTATGCTCTACTATTAACAGACACCGAGCTTCAACGGAAGTTTGTGTTACCAAG 1140
11 111111 111111 111111 111111 111111
1943 A 1943
```









Db 2454 GATCCCTTCATTTACAGACAGACAGAGAGTGTGTCATAGACACAGAGAGGAGTTCCCAAG 2395  
Qy 2002 AAAATTCATTTTAAATCCAGAT 2023  
Db 2394 AACATTCATTTCAACCCGAGGT 2373

Search completed: January 7, 2003, 19:04:12  
Job time : 9564 secs



## RESULT 1

Q9H720 PRELIMINARY; PRT: 699 AA.  
 ID Q9H720  
 AC Q9H720;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
 DE CDNA: FLJ21511 fis, clone COL05748.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=COLON;  
 RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,  
 RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,  
 RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;  
 RT "NEDO human cDNA sequencing project."  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AK025164; BAB15080.1;  
 SQ SEQUENCE 699 AA; 78565 MW; 2F839563189A0523 CRC64;

Query Match 99.8%; Score 3713; DB 4; Length 699;  
 Best Local Similarity 99.9%; Pred. No. 3.6e-263;  
 Matches 698; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MTSWREILLESLLGCVSWSLYHDLGPMIYFPLQTLLELTGLEGFSIAFLSPIFLTITPF 60  
 Db 1 MTSWREILLESLLGCVSWSLYHDLGPMIYFPLQTLLELTGLEGFSIAFLSPIFLTITPF 60  
 Qy 61 WKLNVKKWMLTLLRIITIGSIASFQAPNAKRLMVLALGVSSSLIVQAVTWWSGSHLQRY 120  
 Db 61 WKLNVKKWMLTLLRIITIGSIASFQAPNAKRLMVLALGVSSSLIVQAVTWWSGSHLQRY 120  
 Qy 121 LRIWGFILGQIVLVVLRWIYTSLNPIWSYQMSNKVILTLSAIATLDRICTDGDGCSKPEEK 180  
 Db 121 LRIWGFILGQIVLVVLRWIYTSLNPIWSYQMSNKVILTLSAIATLDRICTDGDGCSKPEEK 180

Qy 181 KTGEVATGMASRPNLLAGAAFGLVFLTHWVFGVSVLSRWAVSGHPHPGDPNPFGGA 240  
 Db 181 KTGEVATGMASRPNLLAGAAFGLVFLTHWVFGVSVLSRWAVSGHPHPGDPNPFGGA 240  
 Qy 241 VLLCLASGLMLPSCLWFRGTGLIWWVTGTASAAGLLYLHTWAAVSGCVFAIFTASMWPQ 300  
 Db 241 VLLCLASGLMLPSCLWFRGTGLIWWVTGTASAAGLLYLHTWAAVSGCVFAIFTASMWPQ 300  
 Qy 301 TLGHLINSGTNPCKTMTIAMIFYLLEIFFCAWCTAFKFPVGGVYARERSDVLLGTMMII 360  
 Db 301 TLGHLINSGTNPCKTMTIAMIFYLLEIFFCAWCTAFKFPVGGVYARERSDVLLGTMMII 360  
 Qy 361 GLNMLFGPKKNLDDLQTKNSSKVLFRKSEKYMKFLWLLVGVGLLGLLRHKAYERKLG 420  
 Db 361 GLNMLFGPKKNLDDLQTKNSSKVLFRKSEKYMKFLWLLVGVGLLGLLRHKAYERKLG 420  
 Qy 421 KVAPTKEVSAAIWPFREFGYDNEGWSLERSAHLNETGADFITILESDASKPYMGNDLT 480  
 Db 421 KVAPTKEVSAAIWPFREFGYDNEGWSLERSAHLNETGADFITILESDASKPYMGNDLT 480  
 Qy 481 MWLGEKLGFTDFGPSTRYHTWGIMALSRYPIVKSEHLLPSPEGEIAPAITLTVNISGK 540  
 Db 481 MWLGEKLGFTDFGPSTRYHTWGIMALSRYPIVKSEHLLPSPEGEIAPAITLTVNISGK 540  
 Qy 541 LVDFVVTTHFGNHEDDLDRKLQAIKSKLLKSSSNQVIFLGYITSAPGSRDYQLTEHGNV 600  
 Db 541 LVDFVVTTHFGNHEDDLDRKLQAIKSKLLKSSSNQVIFLGYITSAPGSRDYQLTEHGNV 600  
 Qy 601 KDIDSTDHWRCEYIMYRGLIRLGYARISHAELSDSEIQMAKFRIPDDPTNYRDNQKVVI 660  
 Db 601 KDIDSTDHWRCEYIMYRGLIRLGYARISHAELSDSEIQMAKFRIPDDPTNYRDNQKVVI 660  
 Qy 661 DHREVSEKIHFNPRFGSYKEGHNYENNHHFMNTPKYFL 699  
 Db 661 DHREVSEKIHFNPRFGSYKEGHNYENNHHFMNTPKYFL 699

*deparagine*

**THIS PAGE BLANK (USPTO)**

RESULT 1  
AK025164 2486 bp mRNA linear PRI 29-SEP-2000  
OCUS Homo sapiens cDNA: FLJ21511 fis, COL05748.  
DEFINITION  
ACCESSION AK025164  
VERSION AK025164.1 GI:10437625  
KEYWORDS oligo capping; fis (full insert sequence).  
SOURCE Homo sapiens colon cDNA to mRNA, clone\_lib:COL clone:COL05748.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (sites)  
AUTHORS Kawabata,A., Hikiiji,T., Kobatake,N., Inagaki,H., Ikema,Y.,  
Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,  
Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.

TITLE NEDO human cDNA sequencing project  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 2486)  
AUTHORS Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,  
Shibahara,T., Tanaka,T. and Nakamura,Y.

TITLE Direct Submission  
JOURNAL Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science,  
University of Tokyo, Laboratory of Genome Structure Analysis, Human  
Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639,  
Japan (E-mail:cdnal@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,  
Fax:81-3-5449-5416)

COMMENT NEDO human cDNA sequencing project supported by Ministry of  
International Trade and Industry of Japan; cDNA full insert  
sequencing: Research Association for Biotechnology; cDNA library  
construction, 5' & 3'-end one pass sequencing: Department of  
Virology and Human Genome Center, Institute of Medical Science,  
University of Tokyo (partly supported by Science and Technology  
Agency).

FEATURES  
source Location/Qualifiers  
1..2486  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="COL05748"  
/tissue\_type="colon"  
/clone\_lib="COL"  
/note="cloning vector pME18SFL3"  
CDS  
183..2282  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="BAB15080.1"  
/db\_xref="GI:10437626"  
/translation="MTSLWREILLESLLGCVSWSLYHDLGPMIYYFPLQTELTGLEG  
FSIAFLSPIFLTITPFWKLVNKKWMLTLLRIITIGSIASFQAPNAKLRMLVLAGVSS  
SLIVQAVTWGSGSLQRYLRIWGFILGQIVLVLRWYTSLNPIWSYQMSNKKVILTSL  
AIATLDRICTDGDGCKPEEKKTGEVATGMASRPNWLACAAFGSLVFLTHWVFEVSL  
VSRWAVSGHPPHGPDPNPFPGAVLLCLASGLMLPSCLEWFRGTGLIWWVTGTASAAGLL  
YLHTWAAVSGCVFAIFTASWNPOTLGHINSNTNPGKMTIAMIFYLLEIFFCAWCT  
AFKEVPGGVYARERSDVLGTMMLIIGLNMLFGPKKNLDLQLTKNSSKVLFRKSEKY  
MKLFELWLVGVLGLGLRHKAYERKLGKVAPTKEVSAAIWPFREFGYDNEGWSLERS  
ALLNETGADFTILESDASKPYMGNDLTMWLGEKLGFTYDFGPSTRYHTWGIMALS  
RYPVIVKSEHLLPSPGEIAPAITLTVNI SGKLVDFVVTFGNHEDDLDRKLQAIASV  
KLLKSSSNQVIFLGYITSAPGSRDYLQLTEHGNVKDIDSTDHWRWCEYIMYRGLIRLG  
YARISHAELSDSEIQMAKFRI PDDPTNYRDNQKVVIDHREVSEKIHFNPRFGSYKEGH  
NYENNHNFMNTPKYFL"

BASE COUNT 692 a 516 c 606 g 672 t  
ORIGIN

Query Match 99.8%; Score 2098.2; DB 9; Length 2486;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 2100; Conservative .0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 ATGACCTCGCTGTGGAGAGAAATCCTCTTGGAGTCGCTGCTGGGATGTGTTCTTGGTCT 60  
|||||  
db 183 ATGACCTCGCTGTGGAGAGAAATCCTCTTGGAGTCGCTGCTGGGATGTGTTCTTGGTCT 242  
Oy 61 CTCTACCATGACCTGGGACCGATGATCTATTACTTTCTTTCGAAACACTAGAACTCACT 120  
|||||  
db 243 CTCTACCATGACCTGGGACCGATGATCTATTACTTTCTTTCGAAACACTAGAACTCACT 302  
Oy 121 GGGCTTGAAGGTTTGTAGTATAGCATTTCTTTCTCCAATATTCCTAACAATTACTCCTTTC 180  
|||||  
db 303 GGGCTTGAAGGTTTGTAGTATAGCATTTCTTTCTCCAATATTCCTAACAATTACTCCTTTC 362  
Oy 181 TGGAAATTGGTTAACAAGAAGTGGATGCTAACCCTGCTGAGGATAATCACTATTGGCAGC 240  
|||||  
db 363 TGGAAATTGGTTAACAAGAAGTGGATGCTAACCCTGCTGAGGATAATCACTATTGGCAGC 422  
Oy 241 ATAGCCTCCTTCCAGGCTCCAAATGCCAAACTTCGACTGATGGTCTTGCCTTGGGGTG 300  
|||||  
db 423 ATAGCCTCCTTCCAGGCTCCAAATGCCAAACTTCGACTGATGGTCTTGCCTTGGGGTG 482  
Oy 301 TCTTCTCTCACTGATAGTGAAGCTGTGACTTGGTGGTGGGAAGTCAATTGCAAGGTAC 360  
|||||

09:15

**THIS PAGE BLANK (USPTO)**

Db 483 TCTTCCTCACTGATAGTGCAAGCTGTGACTTGGTGGTCAGGAAGTCATTGCAAGGTAC 542  
Qy 361 CTCAGAATTTGGGGATTCAATTTTAGGACAGATTGTTCTTGTGTTCTACGCATATGGTAT 420  
Db 543 CTCAGAATTTGGGGATTCAATTTTAGGACAGATTGTTCTTGTGTTCTACGCATATGGTAT 602  
Qy 421 ACTTCACTAAACCAATCTGGAGTTATCAGATGTCCAACAAAGTGATACTGACATTAAGT 480  
Db 603 ACTTCACTAAACCAATCTGGAGTTATCAGATGTCCAACAAAGTGATACTGACATTAAGT 662  
Qy 481 GCCATAGCCACACTTGATCGTATTGGCACAGATGGTGACTGCAGTAAACCTGAAGAAAAG 540  
Db 663 GCCATAGCCACACTTGATCGTATTGGCACAGATGGTGACTGCAGTAAACCTGAAGAAAAG 722  
Qy 541 AAGACTGGTGAGGTAGCCACGGGGATGGCCTCTAGACCCAACCTGGCTGCTGGCAGGGGCT 600  
Db 723 AAGACTGGTGAGGTAGCCACGGGGATGGCCTCTAGACCCAACCTGGCTGCTGGCAGGGGCT 782  
Qy 601 GCTTTTGGTAGCCTTGTTCTCACCCTGGGTTTTTGGAGAAGTCTCTCTTGTGTTCC 660  
Db 783 GCTTTTGGTAGCCTTGTTCTCACCCTGGGTTTTTGGAGAAGTCTCTCTTGTGTTCC 842  
Qy 661 AGATGGGCAGTGAGTGGGCATCCACATCCAGGGCCAGATCCTAACCCTTTGGAGGTGCA 720  
Db 843 AGATGGGCAGTGAGTGGGCATCCACATCCAGGGCCAGATCCTAACCCTTTGGAGGTGCA 902  
Qy 721 GTACTGCTGTGCTTGGCAAGTGGATTGATGCTTCCATCTTGTGTTGCTGCTGGTACT 780  
Db 903 GTACTGCTGTGCTTGGCAAGTGGATTGATGCTTCCATCTTGTGTTGCTGCTGGTACT 962  
Qy 781 GGTGTTGATCTGGTGGGTTACAGGAACAGCTTCAGCTGCGGGGCTCCTTTACCTGCACACA 840  
Db 963 GGTGTTGATCTGGTGGGTTACAGGAACAGCTTCAGCTGCGGGGCTCCTTTACCTGCACACA 1022  
Qy 841 TGGGCAGCTGCTGTGTCTGGCTGTGCTTTCGCCATCTTACTGCATCCATGTGGCCCCAA 900  
Db 1023 TGGGCAGCTGCTGTGTCTGGCTGTGCTTTCGCCATCTTACTGCATCCATGTGGCCCCAA 1082  
Qy 901 ACACTTGGACACCTTATTAACCTCAGGGACAAACCTGGGAAAACCATGACCATTGCCATG 960  
Db 1083 ACACTTGGACACCTTATTAACCTCAGGGACAAACCTGGGAAAACCATGACCATTGCCATG 1142  
Qy 961 ATATTTTATCTTCTAGAAATATTTTCTGTGCCTGGTGCACAGCTTTTAAGTTGTGCCA 1020  
Db 1143 ATATTTTATCTTCTAGAAATATTTTCTGTGCCTGGTGCACAGCTTTTAAGTTGTGCCA 1202  
Qy 1021 GGAGGTGTCTACGCTAGAGAAAGATCAGATGTGCTTTTGGGGACAATGATGTTAATTATC 1080  
Db 1203 GGAGGTGTCTACGCTAGAGAAAGATCAGATGTGCTTTTGGGGACAATGATGTTAATTATC 1262  
Qy 1081 GGGCTGAATATGCTATTTGGTCCTAAGAAAAACCTTGACTTGCTTCTTCAACAAAAAAC 1140  
Db 1263 GGGCTGAATATGCTATTTGGTCCTAAGAAAAACCTTGATTGCTTCTTCAACAAAAAAC 1322  
Qy 1141 AGTTCTAAAGTGCTTTTCAGAAAGAGTGAAAAATACATGAAACTTTTCTGTGGCTGCTT 1200  
Db 1323 AGTTCTAAAGTGCTTTTCAGAAAGAGTGAAAAATACATGAAACTTTTCTGTGGCTGCTT 1382  
Qy 1201 GTTGGTGTGGGATTGTTGGGATTAGGACTACGGCATAAAGCCTATGAGAGAAAACCTGGGC 1260  
Db 1383 GTTGGTGTGGGATTGTTGGGATTAGGACTACGGCATAAAGCCTATGAGAGAAAACCTGGGC 1442  
Qy 1261 AAAGTGGCACCAACCAAGAGGTCTCTGCTGCCATCTGGCCTTTCAGGTTTGGATATGAC 1320  
Db 1443 AAAGTGGCACCAACCAAGAGGTCTCTGCTGCCATCTGGCCTTTCAGGTTTGGATATGAC 1502  
Qy 1321 AATGAAGGGTGGTCTAGTCTAGAAAGATCAGCTCACCTGCTCAATGAAACAGGTGCAGAT 1380  
Db 1503 AATGAAGGGTGGTCTAGTCTAGAAAGATCAGCTCACCTGCTCAATGAAACAGGTGCAGAT 1562  
Qy 1381 TTCATAACAATTTTGGAGAGTGATGCTTCTAAGCCCTATATGGGGAACAATGACTTAACC 1440  
Db 1563 TTCATAACAATTTTGGAGAGTGATGCTTCTAAGCCCTATATGGGGAACAATGACTTAACC 1622

**THIS PAGE BLANK (USPTO)**

QY 241 ATAGCCTCC  
Db 423 ATAGCCTCC  
301 TCTTCCTCACTGATAGTC  
09:15:46 2003

us-09-809-

Qy 1441 ATGTGGCTAGGGGAAAAGTTGGGTTTCTATACAGACTTTGGTCCAAGCACAAGGTATCAC 1500  
Db 1623 ATGTGGCTAGGGGAAAAGTTGGGTTTCTATACAGACTTTGGTCCAAGCACAAGGTATCAC 1682  
Qy 1501 ACTTGGGGGATTATGGCTTTGTCAAGATACCCAATTGTGAAATCTGAGCATCACCTTCTT 1560  
Db 1683 ACTTGGGGGATTATGGCTTTGTCAAGATACCCAATTGTGAAATCTGAGCATCACCTTCTT 1742  
Qy 1561 CCGTCACCAGAGGGCGAGATCGCACCAGCCATCACATTGACCGTTAACATTTCTGGGCAAG 1620  
Db 1743 CCGTCACCAGAGGGCGAGATCGCACCAGCCATCACATTGACCGTTAACATTTCTGGGCAAG 1802  
Qy 1621 CTGGTGGATTTTGTCTGACACACTTTGGGAACCCAGAAAGATGACCTCGACAGGAAACTG 1680  
Db 1803 CTGGTGGATTTTGTCTGACACACTTTGGGAACCCAGAAAGATGACCTCGACAGGAAACTG 1862  
Qy 1681 CAGGCTATTGCTGTTTCAAAACTACTGAAAAGTAGCTCTAATCAAGTGATATTTCTGGGA 1740  
Db 1863 CAGGCTATTGCTGTTTCAAAACTACTGAAAAGTAGCTCTAATCAAGTGATATTTCTGGGA 1922  
Qy 1741 TATATCACTTCAGCACCTGGCTCCAGAGATTATCTACAGCTCACTGAACATGGCAATGTG 1800  
Db 1923 TATATCACTTCAGCACCTGGCTCCAGAGATTATCTACAGCTCACTGAACATGGCAATGTG 1982  
Qy 1801 AAGGATATCGACAGCACTGATCATGACAGATGGTGTGAATACATTATGTATCGAGGGCTG 1860  
Db 1983 AAGGATATCGACAGCACTGATCATGACAGATGGTGTGAATACATTATGTATCGAGGGCTG 2042  
Qy 1861 ATCAGGTTGGGTTATGCAAGAATCTCCCATGCTGAAGTGAAGTATTCAGAAATTCAGATG 1920  
Db 2043 ATCAGGTTGGGTTATGCAAGAATCTCCCATGCTGAAGTGAAGTATTCAGAAATTCAGATG 2102  
Qy 1921 GCAAAATTTAGGATCCCTGATGACCCCACTAATTATAGAGACAACCAGAAAGTGGTCATA 1980  
Db 2103 GCAAAATTTAGGATCCCTGATGACCCCACTAATTATAGAGACAACCAGAAAGTGGTCATA 2162  
Qy 1981 GACCACAGAGAAGTTTCTGAGAAAAATTCATTTTAATCCAGATTGGATCCTACAAAGAA 2040  
Db 2163 GACCACAGAGAAGTTTCTGAGAAAAATTCATTTTAATCCAGATTGGATCCTACAAAGAA 2222  
Qy 2041 GGACACAATTATGAAAACAACCATCATTTTCATATGAATACTCCCAAACTTTTATGA 2100  
Db 2223 GGACACAATTATGAAAACAACCATCATTTTCATATGAATACTCCCAAACTTTTATGA 2282  
Qy 2101 AAC 2103  
Db 2283 AAC 2285

**THIS PAGE BLANK (USPTO)**